Tissue and population-level diversity in plant secondary metabolism: a systematic exploration using MS/MS structural analysis

Dissertation
(Kumulativ)

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr.rer.nat.)

vorgelegt

dem Rat der Chemisch-Geowissenschaftlichen Fakultät

der Friedrich-Schiller-Universität Jena

von

M.S., Dapeng Li
geboren am 01.11.1986 in China

MAX-PLANCK-GESELLSCHAFT

Max-Planck-Institut für chemische Ökologie

Gutachter:

1. Prof. Dr. Ian Thomas Baldwin

   Max Planck Institute for Chemical Ecology, Jena, Germany

2. Prof. Dr. Georg Pohnert

   Institute for Inorganic and Analytical Chemistry, Jena, Germany

3. Prof. Dr. Jean-Luc Wolfender

   School of Pharmaceutical Sciences, University of Geneva, Switzerland

Beginn der Promotion: 08. May 2012

Eingereicht am: 26. May 2016

Tag der Verteidigung: 09. November 2016

Table of Contents

1. General Introduction

1.1. Plant secondary metabolism diversity

Plants are amazing synthetic chemists that create an enormous diversity of low molecular weight compounds not directly essential for energy-related and growth processes, referred to as secondary metabolites. It is estimated that there are 200,000 of these chemical structures existing in the plant kingdom and somewhere between 5,000 and 15,000 within each individual plant species (D'Auria and Gershenzon, 2005; Fernie, 2007). Secondary metabolism, compared to primary metabolism, contributes to the majority of this chemical space due to its large structural diversity. Historically, secondary metabolites were long considered as "flotsam and jetsam on the metabolic beach" or as waste products from metabolic degradation (Mothes, 1955; Haslam, 1986). It was not until key discoveries on certain of the ecological roles that secondary metabolites fulfill in nature (Fraenkel, 1959; Fritz and Simms, 1992), that it became more widely accepted that secondary metabolites are essential determinants of plants' adaptation to their environment. Plant secondary metabolites can act for plants' protection to drought and excess UV radiation, as chemical shields against natural enemies, as allelochemicals against competitor plant species or chemical attractants acting on pollinators or seed dispersing animals (Wink, 2003; Ramakrishna and Ravishankar, 2011).

One of the most striking characteristics of plant secondary metabolites is structural chemical diversity. Each secondary metabolite can exhibit strong qualitative variations across genotypes of a given species without disastrous consequences for growth and developmental outcomes. Additionally, structural variations in secondary metabolite composition can be apprehended at different levels including among plant species, populations and individuals. It is easily noticeable that particular secondary metabolites as well as complete metabolic groups are taxonomically distributed. As such, certain secondary metabolite classes have been used as typical characters for biochemical investigation of specific plant families, for example

quinolizidine alkaloids and non-protein amino acids for Fabaceae; tropane and steroidal alkaloids for Solanaceae, and iridoids and essential oils for Lamiaceae (Wink, 2003). This important metabolic diversification suggests that particular taxa-specific compounds may have been selected by natural selection when they address particular ecological needs. Within and among populations of a species, many secondary metabolites vary quantitatively and qualitatively to maintain population fitness in fluctuating and geographically dispersed environments. For instance, phytohormones regulate various aspects of plant growth and defense in response to environmental cues. Variations in jasmonate accumulation capacities dramatically influence a plant's performance in nature, notably by affecting the inducibility of insecticidal secondary metabolites (Kallenbach et al., 2012). Within a plant, quantitative and qualitative secondary metabolite variations occur among different tissue types, according to ontogenic stages, circadian or annual cycles, and during stress adaptation processes (Kim et al., 2011; Holeski et al., 2012; Moore et al., 2014). Young and reproductive tissues often exhibit the highest stress-responsive secondary metabolite profiles and secondary metabolite compositions change dramatically during fruit maturation (Moco et al., 2007). Many secondary metabolites inducible upon elicitation are synthesized *de novo*. When leaves of tobacco plants are mechanically wounded or attacked by a shoot herbivore, the plants activate synthesis of nicotine in the roots which is then mobilized to the shoots to deter their enemies (Baldwin, 1998). The inducible defense mechanism may allow plants to reduce the costs associated with defense when enemies are absent (Karban et al., 1997).

1.2. The "raison d'être" of plant secondary metabolite diversity

Given that plants are incapable of moving away from threats, they employ particular physiological and chemical responses to cope with unfavorable changes in their environment. Plants have notably evolved biosynthetic capacities for a vast array of secondary metabolites that can help them solving some of these ecological challenges. Accumulating evidence has revealed that plant secondary metabolites carry out many important ecological and physiological functions. Allelochemicals such as alkaloids glucosinolates, terpenes, saponins

and tannins represent adaptive traits that have diversified during evolution by natural selection in order to protect against viruses, bacteria, fungi, competing plants and most importantly, against herbivores (Wink, 1993). In addition, many secondary metabolites may concomitantly possess multiple functions: anthocyanins or monoterpenes function as repellent compounds for herbivores and parasites, while also serving as attractants in flowers (Quina et al., 2009); flavonoids in defense against UV-B radiation and pathogen infection, nodulation, and pollen fertility (Ferreyra et al., 2012). If the costly secondary metabolites play multi-function roles, they probably have a high chance of being maintained during natural selection. Secondary metabolism diversity and variability that represent adaptive characters of plants are likely driven by natural selection during evolution where highly variable environmental interactions take place.

Interaction with insects is considered as one of the key selection pressures that has sculpted plant metabolism diversity. As early as 1964, Ehrlich and Raven proposed that a tight "arms race" between plant defense evolution and herbivore counter-adaptation is responsible for the most of the diversity in secondary metabolites detected across plant species (Ehrlich and Raven, 1964). Yet, plant defense comes at a price. The production, storage, translocation and auto toxicity of secondary metabolites are all costly, and plants only have access to limited resources. Optimal defense theory, which assigns to a tissue with a price for its value to the plant and weights the risk of attacks and the cost-benefit ratio of defense production, provides a conceptual framework to test why tissue-level secondary metabolite variation occur in plants (Rhoades and Cates, 1976). All these factors change throughout plant tissues as well as ontogeny, resulting in a diversified production of secondary metabolites (Barton and Koricheva, 2010). The tissue-level investigation of 17-hydroxygeranyllinalool-diterpene glycosides (HGL-DTGs) accumulation is found highly abundant in flower tissues of *N. attenuata,* which correlates to its high reproductive values to the plants. The distribution of nicotine also follows this same trend (Barton and Koricheva, 2010; Heiling et al., 2010). This is consistent with the observations that defensive secondary metabolites are typically much higher in reproductive, younger and developing tissues than old tissues. As well, reproductive

organs such as flowers usually have stronger defense response comparing with their vegetative counterparts.

Jones & Firn (1991) proposed another simple model to answer the question of "why" there exists high structural diversity in plant secondary metabolism: the screening hypothesis (Jones and Firn, 1991). In the screening hypothesis, they suggested that, due to the low chance of producing bioactive compounds from mutation, plants are more likely to maintain a diverse and rapidly mutating complex blend of secondary metabolites to increase their probability of generating new "active" compounds. As a result, plants are producing a vast array of secondary metabolites for "screening" their biological activity. However, the large number of "inactive" metabolites is unlikely to be retained during natural selection due to their costly production maintenance. A recent study on secondary metabolite evolution in response to changing herbivory showed that polymorphism in a glucosinolate-related locus in *A. thaliana* populations correlates with the distribution of two populations of specialist aphids and that this correlation pattern can be rapidly re-established within only five generations (Züst et al., 2012). Pyrrolizidine alkaloids were shown to be frequently lost and re-gained throughout the evolution of *Senecio* based on biotic stresses faced by these populations (Pelser et al., 2005). These observations suggest that local abiotic and biotic selection pressures are probably the main drivers of secondary metabolism diversity assemblies.

From a mechanistic point of view, gene duplication has been recognized as a major force in diversifying secondary metabolite biosynthesis pathways (Kliebenstein et al., 2001; Kliebenstein et al., 2001; Kliebenstein, 2008). When a gene encoding an enzyme is duplicated, the gene copies are liberated from functional constraints allowing one copy to maintain its original function while the other copy evolves a new product profile leading to neofunctionalization (Lynch and Force, 2000). Genes encoding secondary metabolism-related enzymes often evolved from genes associated with primary metabolism (Schuster et al., 2006). In contrast to primary metabolism, in which selection constrained mutations to maintain the most stable and functional enzyme forms to perform stable outputs, the regulation of secondary metabolism is often underlined by small series of biosynthetic-related genes

(Kliebenstein et al., 2001). This enzyme promiscuity which allows few enzymes to catalyze multiple reactions and biosynthesize multiple products is the mechanistic basis and the prerequisite for the diversification of secondary metabolism and its adaption to the continuously changing environment (Weng et al., 2012).

1.3. Population-level variations in secondary metabolism engaged by plant-insect interactions

Plant species are characterized by a unique blend of secondary metabolites; these mixtures frequently differ in their concentrations and compositions within and among populations of the species (Bravo-Monzon et al., 2014). Population-level variations in secondary metabolism can appear in the form of patterns of presence or absence for individual secondary metabolite, but more frequently appear as quantitative variations (Moore et al., 2014). Quantitative variations among populations are often attributable to the effects of abiotic environmental factors such as temperature, altitude and latitude (Binns et al., 2002; Ganzera et al., 2008), and most importantly, to the consequences of interactions with biotic factors such as insects. The defensive function of secondary metabolites against herbivory and pathogens is considered to be a major driving force of the maintenance of metabolite diversity through evolution (Ehrlich and Raven, 1964). In plants, the pattern of phenotypic variation in secondary metabolism has strong ecological significance and plays a major role in plant–insect interactions by determining plant susceptibility or resistance to herbivores. Variations in secondary metabolism in natural populations may be therefore an indicator of current or historical patterns of interactions with local herbivore community. Moreover, plant defense and growth trade-offs likely provide another important selection pressures that maintain the extensive metabolic polymorphisms commonly observed in natural populations.

The study of natural variation has profoundly advanced our understanding of plants' phenotypic trait evolution. Natural variation of metabolic traits has been long used in quantitative genetics ever since the first application of quantitative trait loci (QTLs) to

inferring the metabolic control of seed colors in *Phaseolus vulgaris* (Sax, 1923). Ecologists, on the other hand, have traditionally focused on how plants face particular ecological pressures and analyzed variations in certain metabolic classes (Strauss and Agrawal, 1999; Kuzina et al., 2011). Current advances in genomics and metabolomics have enabled biologists to explore natural variation in a broader context. This includes the quantification of thousands of metabolites in hundreds of genotypes as well as the concurrent measurement of plant transcriptomes (Fiehn et al., 2008). These advances provide the opportunity to query the causal genes underlying metabolic variations as shown in both model and non-model plant species (Fu and Xue, 2010; Matsuba et al., 2013; Li et al., 2015). However, the analysis of intraspecific variations in metabolism has frequently been biased towards central metabolism and its impact on energy and growth processes or to secondary metabolite classes, for which a priori knowledge exists regarding their biological function (Keurentjes and Sulpice, 2009; Carreno-Quintero et al., 2012). A systematic exploration of qualitative and quantitative variations of metabolic adjustments across native populations in response to herbivory is therefore expected to provide broad insights into secondary metabolism variations and the underlying complex traits and factors that maintain these variations in natural populations. In **manuscript I**, I explored natural diversity patterns in the metabolic response *to Manduca sexta* herbivory of different *Nicotiana attenuata* populations. I employed a computational pipeline to assemble MS/MS data collected in a non-targeted manner and established mass spectral maps using a bioinformatics method to visualize metabolic branch-specific natural variation effects and to annotate metabolites of interest. In **manuscript II**, I review existing results on population-level and within-plant heterogeneity in jasmonate metabolism and signaling in the context of plant-herbivory interactions and discuss tools to manipulate tissue-specific metabolic sinks in jasmonate signaling as well as new perspectives in unraveling additional herbivory-inducible metabolic layers.

1.4. Tissue-specific (within-plant) metabolic specialization

Within a plant, diversity of secondary metabolites is clearly apparent among different

tissue types and is controlled by the tissue-specific localization of particular biosynthesis pathways. Most secondary metabolites vary in composition as well as in concentration in a spatio-temporal manner for a given plant genotype. For example, blends of secondary metabolites in floral organs greatly differ from those of vegetative tissues (Matsuda et al., 2010). Some metabolites are sometimes exclusively present in specific organs in which they fulfill specific biological or physiological functions. 6-tuliposide B, for instance, a secondary metabolite with antibacterial activity accumulates specifically in tulip anthers, but not other tuliposide congeners, such as 6-tuliposide A, with functions as insecticides but also causing repellent to pollinators (Shoji et al., 2005; Shigetomi et al., 2010; Nomura et al., 2015). Some tissues or cell types are also extremely distinguishable from others in terms of secondary metabolite profile. Glandular-secreting trichomes, in particular, are chemically fascinating cells as they are equipped with efficient metabolic systems to produce highly specialized metabolites that fulfill important functions for a plant's interaction with its environment (Weinhold and Baldwin, 2011). Comparisons of glucosinolates in different *Arabidopsis thaliana* organs revealed that plants have evolved tissue-specific metabolic systems to increase their fitness (Brown et al., 2003).

Exploration of tissue-level metabolic specialization is of high relevance for understanding the metabolic contribution for a given tissue to the organism fitness. However, tissue-level exploration of secondary metabolites, except that of the AtMetExpress database (Matsuda et al., 2010), is often biased towards specific groups of secondary metabolites, such as glucosinolates in *Arabidopsis*, or to primary metabolites as a whole (Brown et al., 2003; Desbrosses et al., 2005). In **manuscript III**, I employed a computational pipeline combining tissue-wide non-targeted MS/MS data acquisition and information theory on 14 dissected tissues of *N. attenuata* to mine patterns of tissue-specific structural diversity of secondary metabolites. The resulting MS/MS molecular networks facilitates the exploration of metabolites contributing most to tissue-level specialization.

1.5. Genomic and regulatory bases to the plasticity of plant metabolic responses

The high degree of plasticity of secondary metabolism pathways is supported by the existence in plant genomes of large gene families that encode for metabolic enzymes. Genome sequencing and gene annotation have revealed the existence of large gene families participating in secondary metabolism (metabolism-related genes) such as terpene synthases, cytochrome P450 monooxygenases, UDP-glycosyltransferases, acyltransferases or *O*-methyltransferases (D'Auria and Gershenzon, 2005). Much of secondary metabolite diversity comes from modifications to the core structures of secondary metabolites by metabolism-related genes encoded enzymes. For example, the genome of *Arabidopsis* contains 272 genes belonging to the cytochrome P450 superfamily, 120 UDP-dependent glycosyltransferase (Bowles, 2002; Schuler and Werck-Reichhart, 2003), as well as at least four families of acyltransferases with the BAHD acyltransferase family having 64 members (D'Auria et al., 2002) and the family of serine carboxypeptidase-like acyltransferases having 53 members (Milkowski and Strack, 2004).

In each of these metabolism-related gene families, the vast majority of the genes remain uncharacterized. This is either because the metabolites that they produce are unknown or no significant association has yet been identified between the expressions of specific metabolic genes and characterized metabolic groups or phenotypes of interest. Given the large set of metabolism-related gene families and our scarce knowledge on secondary metabolism in plants, the functional characterization of metabolic genes has remained very challenging.

1.6. *Nicotiana attenuata*: a model system for secondary metabolism mediated ecological interactions

Nicotiana attenuata, an annual wild tobacco species native to the Great Basin Desert in the United States, primarily occurs in large ephemeral populations in post-fire habitats and smaller persistent populations found in washes (Bahulikar et al., 2004). Dormant seeds of this species germinate from long-lived seed banks in sagebrush and pinyon-juniper ecosystems

when fires pyrolize the litter layer, removing germination inhibitors and saturating the soils with smoke-derived germination cues (Baldwin and Morse, 1994; Baldwin et al., 1994). This special fire-synchronized germination behavior makes of *N. attenuata* one of the primary food source for herbivores from various feeding guilds, for example, grasshoppers, leafhoppers, lepidopteran larvae, beetles and stem-boring weevils, that colonize the ecosystem after fires (Kessler et al., 2004; Diezel et al., 2011; Gaquerel et al., 2013). On the other hand, this post-fire germination behavior also creates an intense selection for rapid growth with conspecifics and intense competition for resources leading to exacerbated plant's growth-defense trade-offs. Interaction with insects for pollination services stands out as another important ecological factors that shape secondary metabolite diversity. The synchronized germination behavior on burned habitats translates into patchy *N. attenuata* monocultures that need to lure in pollinators from long distances. In addition, abiotic stresses such as high UV radiation and drought in the Great Basin Desert represent other crucial environmental factors to which *N. attenuata* is shaped to adapt. The complex ecological interactions have therefore sculpted *N. attenuata* diversified secondary metabolism for its survival in nature. From a genetics aspect, the particular germination behavior of *N. attenuata* affects the genetic structure of ephemeral monocultures produced by this species and results in relatively high within-population variation, thought to be in part due to the recruitment of plants of very different ages into post-fire populations from the long-lived seed banks (Bahulikar et al., 2004). This high genetic variations existing in *N. attenuata* populations likely explains the great level of quantitative natural variation in secondary metabolism detected in this species (Li et al., 2015). The *N. attenuata* model system is shown in **Figure 1**.

Figure 1. The *Nicotiana attenuata* model system: photographs from *N. attenuata*'s native habitat in the Great Basin Desert of southwestern Utah. (A) a native post-fire population of *N. attenuata* in southwestern Utah. (B) a zoom-in view of native *N. attenuata* population in early elongation. (C) pollinator: *Manduca sexta* moth. (D) pollinator: hummingbird *Archilochus alexandri*. (E) larvae of specialist *Manduca* spp. (F) tobacco budworm *Heliothis virescens*. (G) larvae of generalist *Spodoptera exigua*. (H) the suckfly *Tupiocoris notatus*. (I) *Empoasca* spp.

leafhoppers. (J) the predatory big-eyed bug *Geocoris* sp.

The highly variable environments and the diverse selection pressures in which this plant grows likely resulted in the rich secondary metabolism of this species. In recent years, several efforts have been undertaken to functionally characterize a vast array of genes controlling metabolic resistance traits to native herbivores in this species including nicotine, a neurotoxin that functions synergistically with anti-digestive plant proteins (Steppuhn et al., 2004; Steppuhn and Baldwin, 2007), phenolic derivatives (Kaur et al., 2010; Onkokesung et al., 2012) and 17-hydroxygeranyllinalool diterpene glycosides (HGL-DTGs) (Heiling et al., 2010; Heiling, 2016) that exhibit high tissue-specific responses to herbivory and *O*-acyl sugars that express preferentially in trichomes and function as plant indirect defenses to herbivory (Weinhold and Baldwin, 2011) (**Figure 2**).

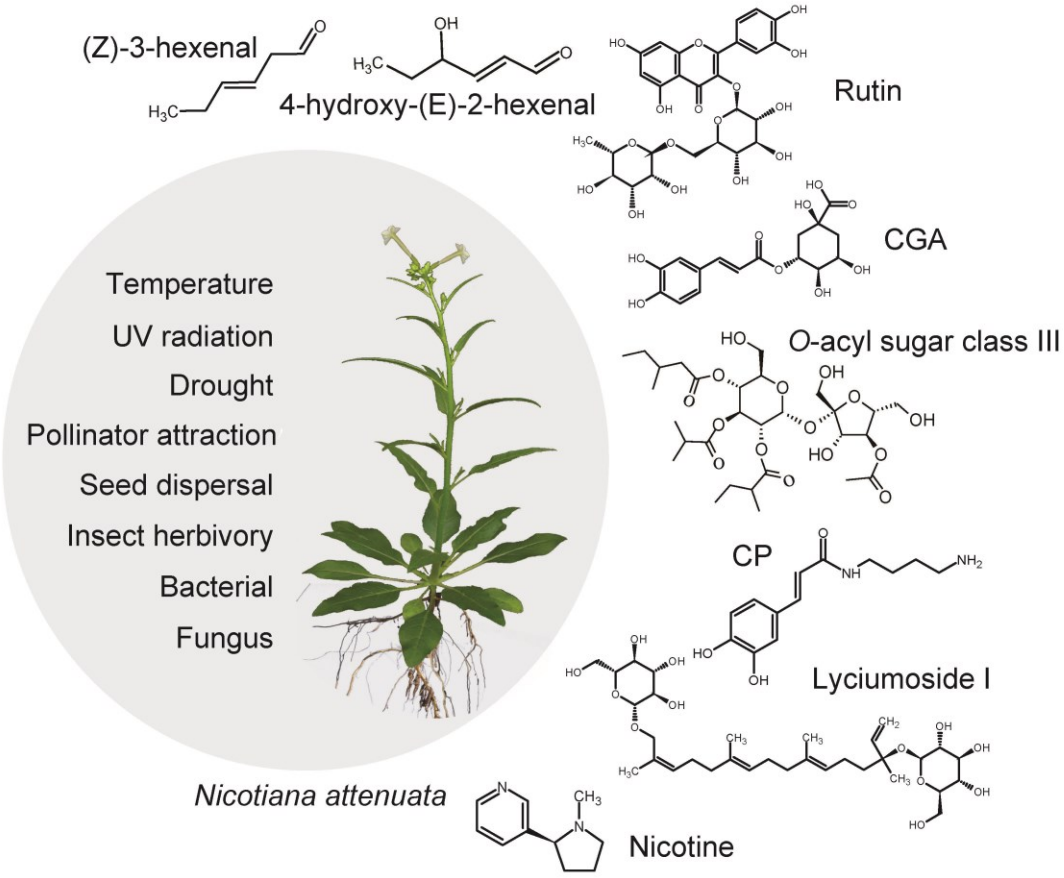

Figure 2. Representative secondary metabolites mediating *Nicotiana attenuata* ecological

interactions. Abiotic and biotic factors influencing the production of these metabolites are shown in the circle. A diverse collection of selected secondary metabolites representing major compound classes in response to stress conditions (both direct and indirect ways) is shown beside the circle. Many of these secondary metabolites vary quantitatively and qualitatively among different tissues and populations. CGA, chlorogenic acid; CP, caffeoylputrescine.

1.7. Challenges and promises in analyzing secondary metabolite diversity using mass spectrometry (MS)-based metabolomics

Compared with transcriptomics approaches that provide almost complete coverage of transcriptome (Gonzalez-Ballester et al., 2010) and proteomics approaches that are now capable of detecting somewhere around 45% of the cellular protein complement (Baerenfaller et al., 2008), metabolomics can only cover a relatively small fraction of the metabolome found in any single cell due to the high structural diversity, complexity of plant metabolites, and their broad dynamic range in cellular abundance (Fernie et al., 2011). Although we have witnessed tremendous advances in the detection and identification of plant metabolites in the last decade using analytical approaches such as liquid chromatography-mass spectrometry (LC-MS) and NMR; the analysis of intraspecific variations in metabolism, however, has lagged behind and frequently been biased towards specific secondary metabolite classes, for which a priori knowledge exists regarding their biological function, or to well-mapped parts of primary metabolism associated with energy and growth processes (Keurentjes and Sulpice, 2009; Carreno-Quintero et al., 2012).

Non-targeted approaches with large-scale tandem mass spectrometry (MS/MS) data acquisition can extensively capture metabolic diversity and informatively infer preliminary structural insights on the metabolites. Indiscriminant (data-independent) MS/MS (idMS/MS) analysis, sometimes referred to as shotgun or broad-scale MS/MS, has gained considerable interest as an exploratory method implemented at the forefront of metabolomics measurements (Broeckling et al., 2013). Data-dependent MS/MS analysis that select several abundant precursor ions for collision induced dissociation (CID) fragmentation, on the other

hand, can sometimes be limited to capture the metabolic diversity in a data-set. IdMS/MS, which fragments all ions that elute from the column indiscriminately, offers many advantages in terms of rapidity and scale of analysis. It is obvious however, that the challenge for idMS/MS is in its computational assignment of precursor-fragment relationships, where information is lost due to the idMS/MS data acquisition; computational pipelines have been established to perform fragment relationship assignments (Broeckling et al., 2013).

The identification of unknown secondary metabolites represents a critical bottleneck for the exploration of secondary metabolism diversity. The non-targeted MS/MS profiling has its potential to bring insights into structure of unknown metabolites. However, it remains challenging in querying the structure of extremely structurally diverse chemicals due to the fact that most plant metabolite public databases are too sparsely populated to provide useful hints on these unknown structures (Neumann and Bocker, 2010). Visualization of such large-scale MS/MS data remains another on-going challenge. Innovative approaches have therefore to be implemented to accelerate the annotation of MS/MS data. One of the successful approaches pioneered by the Dorrestein laboratory involves the creation of MS/MS molecular networks (Watrous et al., 2012). This approach involves the generation of networks in which mass spectra (nodes) from one data-set are clustered using edges whose length is inversely proportional to the pairwise similarity between spectra. This molecular networking, on the one hand, provides structure hints on the unknown metabolites when they share adjacent metabolic space to known ones, on the other hand, it allows the visualization of large metabolomics data-sets with combined view of other biological information linked to a given node (MS/MS spectrum), such as population-level secondary metabolite natural variation or tissue-wise metabolic diversity. These new approaches facilitate the process of formulating structural hypotheses by the combined interpretation of phenotypic information and mass spectrometric signatures but are not replacing traditional compound identification procedure, a task that is traditionally achieved after compound purification and de novo identification by NMR.

1.8. The power of integrating metabolomics with other omics data for gene function analysis

The integration and analysis of different omics data are providing unprecedented insights into the complex architecture and regulatory networks for a given organism (Joyce and Palsson, 2006). Metabolomics when combined with other omics can be extremely powerful in unraveling the underlying regulatory mechanism of secondary metabolite production. The identification of the genetic bases of metabolic traits via quantitative traits locus (QTL) mapping has been extensively carried out using natural variation in both model and crop species (Johal et al., 2008; Chan et al., 2011; Riedelsheimer et al., 2012). Large-scale metabolic quantitative trait loci (mQTL) combined with correlation-based network analysis can powerfully query the genomic regions harboring particular secondary metabolic traits (Alseekh et al., 2015; Toubiana et al., 2015). Such correlational approaches can be progressively expanded to identify metabolites beyond those with already known pathway assignments by correlating their variation with those detected at the level of metabolic genes, such as recently implemented in rice and maize as well as for non-model system plants (Gong et al., 2013; Wen et al., 2014; Agerbirk et al., 2015; Kusano et al., 2015). Such correlational approaches can be extended to any kind of studies. Coexpression analysis using information about gene and secondary metabolite cross-tissue expression patterns has been successfully applied to infer causative genes in secondary metabolism (Hirai et al., 2007; Rajniak et al., 2015).

In **manuscript I**, I used similar approach in the analysis of large-scale defense metabolism by correlating the variation across different genotypes with the aim of identifying naturally variable and jasmonate signaling-dependent metabolic traits engaged by herbivory. In **manuscript III**, I extended the application of variance from natural populations to different tissues within-a-plant. I constructed cross-tissue co-association maps by combining metabolomics and transcriptomics data using Kurtosis filtering. The resulting tissue-metabolite association maps provide predictive power for tissue-level analysis of gene

functions underlying metabolic signatures.

References

Agerbirk N, Olsen CE, Heimes C, Christensen S, Bak S, Hauser TP (2015) Multiple hydroxyphenethyl glucosinolate isomers and their tandem mass spectrometric distinction in a geographically structured polymorphism in the crucifer Barbarea vulgaris. Phytochemistry **115:** 130-142

Alseekh S, Tohge T, Wendenberg R, Scossa F, Omranian N, Li J, Kleessen S, Giavalisco P, Pleban T, Mueller-Roeber B, Zamir D, Nikoloski Z, Fernie AR (2015) Identification and mode of inheritance of quantitative trait loci for decondary metabolite sbundance in tomato. Plant Cell **27:** 485-512

Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W, Baginsky S (2008) Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. Science **320:** 938-941

Bahulikar RA, Stanculescu D, Preston CA, Baldwin IT (2004) ISSR and AFLP analysis of the temporal and spatial population structure of the post-fire annual, *Nicotiana attenuata*, in SW Utah. BMC Ecol **4:** 12

Baldwin IT (1998) Jasmonate-induced responses are costly but benefit plants under attack in native populations. Proceedings of the National Academy of Sciences of the United States of America **95:** 8113-8118

Baldwin IT, Morse L (1994) Up in smoke: II. Germination of *Nicotiana attenuata* in response to smoke-derived cues and nutrients in burned and unburned soils. Journal of Chemical Ecology **20:** 2373-2391

Baldwin IT, Staszakkozinski L, Davidson R (1994) Up in smoke: I. Smoke-derived germination cues for postfire annual, *Nicotiana attenuata* torr. Ex. Watson. Journal of Chemical Ecology **20:** 2345-2371

Barton KE, Koricheva J (2010) The ontogeny of plant defense and herbivory: characterizing general patterns using meta-analysis. American Naturalist **175:** 481-493

Binns SE, Arnason JT, Baum BR (2002) Phytochemical variation within populations of *Echinacea angustifolia* (Asteraceae). Biochemical Systematics and Ecology **30:** 837-854

Bowles D (2002) A multigene family of glycosyltransferases in a model plant, *Arabidopsis thaliana*. Biochemical Society Transactions **30:** 301-306

Bravo-Monzon AE, Rios-Vasquez E, Delgado-Lamas G, Espinosa-Garcia FJ (2014) Chemical diversity among populations of *Mikania micrantha*: geographic mosaic structure and herbivory. Oecologia **174:** 195-203

Broeckling CD, Heuberger AL, Prince JA, Ingelsson E, Prenni JE (2013) Assigning precursor-product ion relationships in indiscriminant MS/MS data from non-targeted metabolite profiling studies. Metabolomics **9:** 33-43

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. Phytochemistry **62:** 471-481

Carreno-Quintero N, Acharjee A, Maliepaard C, Bachem CW, Mumm R, Bouwmeester H, Visser RG, Keurentjes JJ (2012) Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality. Plant Physiol **158:** 1306-1318

Chan EKF, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. PLoS Biol **9**

D'Auria JC, Chen F, Pichersky E (2002) Characterization of an acyltransferase capable of synthesizing benzylbenzoate and other volatile esters in flowers and damaged leaves of *Clarkia breweri*. Plant Physiol **130:** 466-476

D'Auria JC, Gershenzon J (2005) The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. Curr Opin Plant Biol **8:** 308-316

Desbrosses GG, Kopka J, Udvardi MK (2005) Lotus japonicus metabolic profiling: development of gas chromatography-mass spectrometry resources for the study of plant-microbe interactions. Plant Physiol **137:** 1302-1318

Diezel C, Kessler D, Baldwin IT (2011) Pithy protection: *Nicotiana attenuata*'s jasmonic acid-mediated defenses are required to resist stem-boring *Weevil* larvae. Plant Physiol **155:** 1936-1946

Ehrlich PR, Raven PH (1964) Butterflies and plants - a study in coevolution. Evolution **18:** 586-608

Fernie AR (2007) The future of metabolic phytochemistry: Larger numbers of metabolites, higher resolution, greater understanding. Phytochemistry **68:** 2861-2880

Fernie AR, Aharoni A, Willmitzer L, Stitt M, Tohge T, Kopka J, Carroll AJ, Saito K, Fraser PD, DeLuca V (2011) Recommendations for reporting metabolite data. Plant Cell **23:** 2477-2482

Ferreyra MLF, Rius SP, Casati P (2012) Flavonoids: biosynthesis, biological functions, and biotechnological applications. Frontiers in Plant Science **3**

Fiehn O, Wohlgemuth G, Scholz M, Kind T, Lee do Y, Lu Y, Moon S, Nikolau B (2008) Quality control for plant metabolomics: reporting MSI-compliant studies. Plant J **53:** 691-704

Fraenkel GS (1959) The raison d'etre of secondary plant substances; these odd chemicals arose as a means of protecting plants from insects and now guide insects to food. Science **129:** 1466-1470

Fritz RS, Simms EL (1992) Plant resistance to herbivores and pathogens: ecology, evolution, and genetics. University of Chicago Press

Fu FF, Xue HW (2010) Coexpression analysis identifies Rice Starch Regulator1, a rice AP2/EREBP family transcription factor, as a novel rice starch biosynthesis regulator. Plant Physiol **154:** 927-938

Ganzera M, Guggenberger M, Stuppner H, Zidorn C (2008) Altitudinal variation of secondary metabolite profiles in flowering heads of *Matricaria chamomilla* cv. BONA.

Planta Medica **74:** 453-457

Gaquerel E, Stitz M, Kallenbach M, Baldwin IT (2013) Jasmonate signaling in the field, part II: insect-guided characterization of genetic variations in jasmonate-dependent defenses of transgenic and natural *Nicotiana attenuata* populations. Methods Mol Biol **1011:** 97-109

Gong L, Chen W, Gao Y, Liu X, Zhang H, Xu C, Yu S, Zhang Q, Luo J (2013) Genetic analysis of the metabolome exemplified using a rice population. Proc Natl Acad Sci U S A **110:** 20320-20325

Gonzalez-Ballester D, Casero D, Cokus S, Pellegrini M, Merchant SS, Grossman AR (2010) RNA-seq analysis of sulfur-deprived Chlamydomonas cells reveals aspects of acclimation critical for cell survival. Plant Cell **22:** 2058-2084

Haslam E (1986) Secondary metabolism - fact and fiction. Natural Product Reports **3:** 217-249

Heiling S, Khanal, S., Barsch, A., Zurek, G., Baldwin, I. T., Gaquerel, E. (2016) Using the knowns to discover the unknowns: MS-based dereplication uncovers structural diversity in 17-hydroxygeranyllinalool diterpene glycoside defense production in the Solanaceae. The Plant Journal **85:** 561-577

Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, Schneider B, Jassbi AR, Baldwin IT (2010) Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. Plant Cell **22:** 273-292

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proceedings of the National Academy of Sciences of the United States of America **104:** 6478-6483

Holeski LM, Hillstrom ML, Whitham TG, Lindroth RL (2012) Relative importance of

genetic, ontogenetic, induction, and seasonal variation in producing a multivariate defense phenotype in a foundation tree species. Oecologia **170:** 695-707

Johal GS, Balint-Kurti P, Well CF (2008) Mining and harnessing natural variation: a little MAGIC. Crop Science **48:** 2066-2073

Jones CG, Firn RD (1991) On the evolution of plant secondary chemical diversity. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **333:** 273-280

Joyce AR, Palsson BO (2006) The model organism as a system: integrating 'omics' data sets. Nature Reviews Molecular Cell Biology **7:** 198-210

Kallenbach M, Bonaventure G, Gilardoni PA, Wissgott A, Baldwin IT (2012) Empoasca leafhoppers attack wild tobacco plants in a jasmonate-dependent manner and identify jasmonate mutants in natural populations. Proc Natl Acad Sci U S A **109:** E1548-1557

Karban R, Agrawal AA, Mangel M (1997) The benefits of induced defenses against herbivores. Ecology **78:** 1351-1355

Kaur H, Heinzel N, Schottner M, Baldwin IT, Galis I (2010) R2R3-NaMYB8 regulates the accumulation of phenylpropanoid-polyamine conjugates, which are essential for local and systemic defense against insect herbivores in *Nicotiana attenuata*. Plant Physiol **152:** 1731-1747

Kessler A, Halitschke R, Baldwin IT (2004) Silencing the jasmonate cascade: Induced plant defenses and insect populations. Science **305:** 665-668

Keurentjes JJ, Sulpice R (2009) The role of natural variation in dissecting genetic regulation of primary metabolism. Plant Signal Behav **4:** 244-246

Kim SG, Yon F, Gaquerel E, Gulati J, Baldwin IT (2011) Tissue specific diurnal rhythms of metabolites and their regulation during herbivore attack in a native tobacco, *Nicotiana attenuata*. Plos One **6:** e26214

Kliebenstein DJ (2008) A role for gene duplication and natural variation of gene expression in the evolution of metabolism. Plos One **3:** e1838

Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J,

Mitchell-Olds T (2001) Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. Plant Physiol **126:** 811-825

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. Plant Cell **13:** 681-693

Kusano M, Yang Z, Okazaki Y, Nakabayashi R, Fukushima A, Saito K (2015) Using metabolomic approaches to explore chemical diversity in rice. Mol Plant **8:** 58-67

Kuzina V, Nielsen JK, Augustin JM, Torp AM, Bak S, Andersen SB (2011) *Barbarea vulgaris* linkage map and quantitative trait loci for saponins, glucosinolates, hairiness and resistance to the herbivore *Phyllotreta nemorum*. Phytochemistry **72:** 188-198

Li D, Baldwin IT, Gaquerel E (2015) Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. Proc Natl Acad Sci U S A **112:** E4147-4155

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics **154:** 459-473

Matsuba Y, Nguyen TT, Wiegert K, Falara V, Gonzales-Vigil E, Leong B, Schafer P, Kudrna D, Wing RA, Bolger AM, Usadel B, Tissier A, Fernie AR, Barry CS, Pichersky E (2013) Evolution of a complex locus for terpene biosynthesis in *solanum*. Plant Cell **25:** 2022-2036

Matsuda F, Hirai MY, Sasaki E, Akiyama K, Yonekura-Sakakibara K, Provart NJ, Sakurai T, Shimada Y, Saito K (2010) AtMetExpress development: a phytochemical atlas of Arabidopsis development. Plant Physiol **152:** 566-578

Milkowski C, Strack D (2004) Serine carboxypeptidase-like acyltransferases. Phytochemistry **65:** 517-524

Moco S, Capanoglu E, Tikunov Y, Bino RJ, Boyacioglu D, Hall RD, Vervoort J, De Vos RCH (2007) Tissue specialization at the metabolite level is perceived during the development of tomato fruit. Journal of Experimental Botany **58:** 4131-4146

Moore BD, Andrew RL, Kulheim C, Foley WJ (2014) Explaining intraspecific diversity in plant secondary metabolites in an ecological context. New Phytol **201:** 733-750

Mothes K (1955) Physiology of alkaloids. Annual Review of Plant Physiology and Plant Molecular Biology **6:** 393-432

Neumann S, Bocker S (2010) Computational mass spectrometry for metabolomics: Identification of metabolites and small molecules. Analytical and Bioanalytical Chemistry **398:** 2779-2788

Nomura T, Murase T, Ogita S, Kato Y (2015) Molecular identification of tuliposide B-converting enzyme: a lactone-forming carboxylesterase from the pollen of tulip. Plant Journal **83:** 252-262

Onkokesung N, Gaquerel E, Kotkar H, Kaur H, Baldwin IT, Galis I (2012) MYB8 controls inducible phenolamide levels by activating three novel hydroxycinnamoyl-coenzyme A:polyamine transferases in *Nicotiana attenuata*. Plant Physiol **158:** 389-407

Pelser PB, de Vos H, Theuring C, Beuerle T, Vrieling K, Hartmann T (2005) Frequent gain and loss of pyrrolizidine alkaloids in the evolution of *Senecio* section *Jacobaea* (Asteraceae). Phytochemistry **66:** 1285-1295

Quina FH, Moreira PF, Vautier-Giongo C, Rettori D, Rodrigues RF, Freitas AA, Silva PF, Macanita AL (2009) Photochemistry of anthocyanins and their biological role in plant tissues. Pure and Applied Chemistry **81:** 1687-1694

Rajniak J, Barco B, Clay NK, Sattely ES (2015) A new cyanogenic metabolite in *Arabidopsis* required for inducible pathogen defence. Nature **525:** 376-+

Ramakrishna A, Ravishankar GA (2011) Influence of abiotic stress signals on secondary metabolites in plants. Plant Signal Behav **6:** 1720-1731

Rhoades DF, Cates RG (1976) Toward a general theory of plant antiherbivore chemistry. Recent Advances in Phytochemistry **10:** 168–213

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genome-wide association mapping of

leaf metabolic profiles for dissecting complex traits in maize. Proceedings of the National Academy of Sciences of the United States of America **109:** 8872-8877

Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *PHASEOLUS VULGARIS*. Genetics **8:** 552-560

Schuler MA, Werck-Reichhart D (2003) Functional genomics of P450s. Annual Review of Plant Biology **54:** 629-667

Schuster J, Knill T, Reichelt M, Gershenzon J, Binder S (2006) BRANCHED-CHAIN AMINOTRANSFERASE4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. Plant Cell **18:** 2664-2679

Shigetomi K, Shoji K, Mitsuhashi S, Ubukata M (2010) The antibacterial properties of 6-tuliposide B. Synthesis of 6-tuliposide B analogues and structure-activity relationship. Phytochemistry **71:** 312-324

Shoji K, Ubukata M, Momonoi K, Tsuji T, Morimatsu T (2005) Anther-specific production of antimicrobial tuliposide B in tulips. Journal of the Japanese Society for Horticultural Science **74:** 469-475

Steppuhn A, Baldwin IT (2007) Resistance management in a native plant: nicotine prevents herbivores from compensating for plant protease inhibitors. Ecol Lett **10:** 499-511

Steppuhn A, Gase K, Krock B, Halitschke R, Baldwin IT (2004) Nicotine's defensive function in nature. PLoS Biol **2:** 1074-1080

Strauss SY, Agrawal AA (1999) The ecology and evolution of plant tolerance to herbivory. Trends Ecol Evol **14:** 179-185

Toubiana D, Batushansky A, Tzfadia O, Scossa F, Khan A, Barak S, Zamir D, Fernie AR, Nikoloski Z, Fait A (2015) Combined correlation-based network and mQTL analyses efficiently identified loci for branched-chain amino acid, serine to threonine, and proline metabolism in tomato seeds. Plant Journal **81:** 121-133

Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, van der Voort M, Pogliano K, Gross H, Raaijmakers JM, Moore BS, Laskin J, Bandeira N, Dorrestein PC (2012) Mass spectral molecular networking of living microbial

colonies. Proceedings of the National Academy of Sciences of the United States of America **109:** E1743-E1752

Weinhold A, Baldwin IT (2011) Trichome-derived *O*-acyl sugars are a first meal for caterpillars that tags them for predation. Proceedings of the National Academy of Sciences of the United States of America **108:** 7855-7859

Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J, Yan J (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat Commun **5:** 3438

Weng JK, Philippe RN, Noel JP (2012) The rise of chemodiversity in plants. Science **336:** 1667-1670

Wink M (1993) Allelochemical properties or the raison d'être of alkaloids. The Alkaloids **43:** 1-118

Wink M (2003) Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry **64:** 3-19

Züst T, Heichinger C, Grossniklaus U, Harrington R, Kliebenstein DJ, Turnbull LA (2012) Natural enemies drive geographic variation in plant defenses. Science **338:** 116-119

2. Overview of Manuscripts

Manuscript I

Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis

Dapeng Li, Ian T. Baldwin, and Emmanuel Gaquerel

Published in Proceedings of the National Academy of Sciences of the United States of America 2015, 112(30), E4147-E4155.

The study of natural variation has profoundly advanced our understanding of plants' phenotypic trait evolution. The analysis of intraspecific variations in metabolism, however, has lagged behind and frequently been biased toward central metabolism. To redress this bias, we present a metabolomics case study of leaf secondary metabolites of wild tobacco ecotypes subjected to simulated insect herbivory in which mass spectral maps are constructed. Navigating these maps revealed metabolic branch-specific variations and allowed the annotation of unknown metabolites of likely ecological importance. Although the profiling of entire plant metabolomes remains technically challenging due to their structural complexity, we predict that the workflow described here provides a significant advance in our ability to rapidly explore small molecules that mediate functionally important phenotypes.

Dapeng Li designed the research, performed the research, analyzed data and wrote the paper; Ian T. Baldwin designed the research, contributed new reagents/analytic tools and wrote the paper; Emmanuel Gaquerel designed the research, performed the research, analyzed data and wrote the paper.

Manuscript II

Beyond the canon: within-plant and population-level heterogeneity in jasmonate signaling engaged by plant-insect interactions

Dapeng Li, Ian T. Baldwin, and Emmanuel Gaquerel

Published in Plants 2016, 5(1).

In this manuscript, we provide a short overview of within-plant heterogeneities in jasmonate signaling and dependent responses in the context of plant-insect interactions as illuminated by examples from recent work with the ecological model, *Nicotiana attenuata*. We then discussed means of manipulating jasmonate signaling by creating tissue-specific jasmonate sinks, and the micrografting of different transgenic plants. The metabolic phenotyping of these manipulations provides an integrative understanding of the functional significance of deviations from the canonical model of this hormonal pathway. Additionally, natural variation in jasmonate biosynthesis and signaling both among and within species can explain polymorphisms in resistance to insects in nature. In this respect, insect-guided explorations of population-level variations in jasmonate metabolism have revealed more complexity than previously realized and we discussed how different "omic" techniques can be used to exploit the natural variation that occurs in this important signaling pathway.

Dapeng Li, Ian T. Baldwin, and Emmanuel Gaquerel contributed to the writing, revising and proofreading of the paper.

Manuscript III

Illuminating a plant's tissue-specific metabolic diversity using computational metabolomics and information theory

Dapeng Li, Sven Heiling, Ian T. Baldwin, and Emmanuel Gaquerel

Published in Proceedings of the National Academy of Sciences of the United States of America 2016, 113(47), E7610-E7618.

In this manuscript, we present an exploratory pipeline combining tissue-wide non-targeted mass spectral data acquisition, distribution statistics, information theory and MS/MS molecular networks to the analysis of two different methanolic extracts of 14 dissected tissues of *Nicotiana attenuata* which led to the deconvolution of 895 non-redundant MS/MS spectra. Using an analysis motivated by information theory, we notably detected that anthers harbor the most specialized metabolome and, through the use of MS/MS molecular networks, most unique metabolites to anthers and other tissues were characterized. To refine this view on tissue-level metabolic specialization, we also examined metabolite preferential tissue-associations as well as shared-associations between tissues using Kurtosis filtering. The resulting tissue-metabolite association map provides predictive power for tissue-level analysis of gene functions underlying metabolic signatures, as shown by the gene-silencing-based characterization of two UDP-glycosyltransferases for flavonoid metabolism.

Dapeng Li designed the research, performed the research, analyzed data and wrote the paper; Sven Heiling contributed to virus-induced-gene-silencing experiment and wrote the manuscript; Ian T. Baldwin designed the research, contributed new reagents/analytic tools and wrote the paper; Emmanuel Gaquerel designed the research, performed the research, analyzed data and wrote the paper;

3. Manuscripts

Manuscript I

Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis

Dapeng Li, Ian T. Baldwin, and Emmanuel Gaquerel

Published in Proceedings of the National Academy of Sciences of the United States of America 2015, 112(30), E4147-E4155.

PNAS PLUS

Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis

Dapeng Li[a], Ian T. Baldwin[a], and Emmanuel Gaquerel[a,b,1]

[a]Department of Molecular Ecology, Max Planck Institute for Chemical Ecology, D-07745 Jena, Germany; and [b]Centre for Organismal Studies, University of Heidelberg, 69120 Heidelberg, Germany

Edited by Harry J. Klee, University of Florida, Gainesville, FL, and approved June 11, 2015 (received for review February 13, 2015)

Natural variation can be extremely useful in unraveling the determinants of phenotypic trait evolution but has rarely been analyzed with unbiased metabolic profiling to understand how its effects are organized at the level of biochemical pathways. Native populations of *Nicotiana attenuata*, a wild tobacco species, have been shown to be highly genetically diverse for traits important for their interactions with insects. To resolve the chemodiversity existing in these populations, we developed a metabolomics and computational pipeline to annotate leaf metabolic responses to *Manduca sexta* herbivory. We selected seeds from 43 accessions of different populations from the southwestern United States—including the well-characterized Utah 30th generation inbred accession—and grew 183 plants in the glasshouse for standardized herbivory elicitation. Metabolic profiles were generated from elicited leaves of each plant using a high-throughput ultra HPLC (UHPLC)-quadrupole TOFMS (qTOFMS) method, processed to systematically infer covariation patterns among biochemically related metabolites, as well as unknown ones, and finally assembled to map natural variation. Navigating this map revealed metabolic branch-specific variations that surprisingly only partly overlapped with jasmonate accumulation polymorphisms and deviated from canonical jasmonate signaling. Fragmentation analysis via indiscriminant tandem mass spectrometry (idMS/MS) was conducted with 10 accessions that spanned a large proportion of the variance found in the complete accession dataset, and compound spectra were computationally assembled into spectral similarity networks. The biological information captured by this networking approach facilitates the mining of the mass spectral data of unknowns with high natural variation, as demonstrated by the annotation of a strongly herbivory-inducible phenolic derivative, and can guide pathway analysis.

plant–insect interactions | metabolomics | mass spectrometry | natural variation

Elucidating the structure of metabolites underlying complex traits and the factors that maintain their variation in natural populations are important challenges in plant ecological studies (1). Many studies have notably shown that stress-responsive pathways that produce secondary metabolites are sporadically found across different plant taxa with extensive diversification (2). This important diversification suggests that particular metabolic systems have been recruited through natural selection when the set of compounds that they produce address specific ecological needs. Interactions with insects are important selection pressures that have sculpted plant metabolism, and many plant metabolites protect against herbivore attack and physical damage (3–5). The timely production of particular secondary metabolites in response to insect attack benefits plants by decreasing the costs of constitutive metabolite production. Trade-offs between defense metabolite productions and the intrinsic growth-related functions of central metabolic pathways likely provide important selection pressures that maintain the extensive metabolic polymorphisms commonly observed in natural populations.

Gene discovery strategies exploiting natural variation in quantitative traits, including metabolite levels, have been extensively used in combination with genetic approaches (6–12). Analytical approaches applied in this research field are frequently focused on the quantification of individual or small families of compounds. Procedures such as liquid chromatography-mass spectrometry (LC-MS) and NMR have notably been used with both model and crop species to identify the genetic architecture of metabolic traits using quantitative trait locus mapping approaches (reviewed in ref. 13). Such approaches have been very successful in addressing genomic regions responsible for glucosinolate accumulation in *Arabidopsis* and related species (10, 14–16). Compared with modern sequencing and proteomics technologies, the profiling of entire plant metabolomes is, however, technically unfeasible with the existing analytical platforms, and, as a consequence, the analysis of metabolite natural variation has frequently been biased to secondary metabolite classes, for which a priori knowledge exists regarding their biological function, or to well-mapped parts of primary metabolism associated with energy and growth processes (17–19).

Another critical aspect for exploiting natural variation in metabolism lies in the identification of unknown metabolites that exhibit significant associations with a phenotype of interest. Nontargeted approaches for rapidly collecting repertoires of tandem mass

Significance

The study of natural variation has profoundly advanced our understanding of plants' phenotypic trait evolution. The analysis of intraspecific variations in metabolism, however, has lagged behind and frequently been biased toward central metabolism. To redress this bias, we present a metabolomics case study of leaf secondary metabolites of wild tobacco ecotypes subjected to simulated insect herbivory in which mass spectral maps are constructed. Navigating these maps revealed metabolic branch-specific variations and allowed the annotation of unknown metabolites of likely ecological importance. Although the profiling of entire plant metabolomes remains technically challenging due to their structural complexity, we predict that the workflow described here provides a significant advance in our ability to rapidly explore small molecules that mediate functionally important phenotypes.

Author contributions: D.L., I.T.B., and E.G. designed research; D.L. and E.G. performed research; I.T.B. contributed new reagents/analytic tools; D.L. and E.G. analyzed data; and D.L., I.T.B., and E.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The MS/MS dataset has been deposited in the open metabolomics database Metabolights, www.ebi.ac.uk/metabolights (accession no. MTBLS203).

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: emmanuel.gaquerel@cos.uni-heidelberg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1503106112/-/DCSupplemental.

spectrometry (MS/MS) data can be extremely powerful in capturing the metabolic diversity expected to occur in natural populations (20). Indiscriminant or shotgun MS/MS strategies with high-resolution MS detectors offer many advantages in terms of rapidity and scale of analysis. Pipelines have been recently established to analyze such data (20). However, querying MS/MS data from the analysis of secondary metabolites from public databases is frequently unsuccessful because few standards are available for these compounds (21). An alternative is the use of comparative spectral analysis applied to experimental MS/MS datasets (22). This approach, termed molecular networking, is relatively new and aims at creating a map of mass spectral structural space in which molecules with related MS/MS spectra cluster together. Here, we combine the rapidly generated MS/MS data for all mass signals detected and molecular network construction in the analysis of the metabolic composition of natural plant populations.

We applied our MS method to the natural variation in secondary metabolic profiles observed in accessions of the coyote tobacco, *Nicotiana attenuata*. This annual, native to the Great Basin Desert in the United States, primarily occurs in large ephemeral populations in post-fire habitats and smaller persistent populations found in washes (23). Dormant seeds of this species germinate from long-lived seed banks in sagebrush and pinyon-juniper ecosystems when fires pyrolize the litter layer, removing germination inhibitors and saturating the soils with smoke-derived germination cues (24, 25). This particular germination behavior affects the genetic structure of ephemeral monocultures produced by this species and results in relatively high within-population variation. *N. attenuata* populations represent a primary food source for insects that colonize the ecosystem after fires, and a vast array of genes and dependent metabolic pathways underlying resistance traits to native herbivores have been functionally characterized in this species. Among the major compound classes that contribute to the antiherbivore defense mechanisms of this plant is nicotine, a neurotoxin that functions synergistically with antidigestive plant proteins (26, 27), phenolic derivatives that exhibit strong tissue-specific responses to insect herbivory (28, 29), and 17-hydroxygeranyllinalool diterpene glycosides (HGL-DTGs) (30).

Several studies have analyzed, with a high degree of spatial and temporal resolution, some of the metabolomic reconfigurations that are activated in plant tissues during biotic stresses (for a review, see ref. 31), including the attack of insects (32–34); but few of these studies have explored qualitative and quantitative variations of these metabolic adjustments across native populations. To systematically explore natural diversity patterns in the metabolic response to *Manduca sexta* herbivory of different *N. attenuata* populations, we conducted a glasshouse-based high-throughput MS-based metabolomics approach on 183 plants derived from seeds collected in Utah, Nevada, Arizona, and California. We then optimized an analytical and computational pipeline to assemble MS/MS data collected in a nontargeted manner and established mass spectral maps using a bioinformatics method to visualize metabolic branch-specific natural variation effects and annotate metabolites of interest.

Results and Discussion

***N. attenuata* Populations Exhibit Highly Variable Herbivory-Induced Metabolic Profiles.** To explore patterns of natural variation existing in the herbivory-induced metabolic profile of wild populations of *N. attenuata*, we used a rapid ultra HPLC (UHPLC)-electrospray ionization (ESI)/quadrupole TOFMS (qTOFMS) method to measure the metabolomes of methanolic leaf extracts of 183 individual plants derived from seeds from 43 accessions (*SI Appendix*, Table S1). This analytical procedure allows for the profiling of a broad range of secondary metabolites and their precursors. Seeds used in this experiment had been collected over the last 20 y by Ian T. Baldwin and associates from ephemeral populations growing in postfire habitats in Utah (33 accessions), Nevada (7 accessions), Arizona (2 accessions), and California (1 accession) (Fig. 1*A*). These natural accessions have defined plants surviving as distinct groups through environmental selection and isolation. Our large-scale experiment also included 13 plants from a Utah (U30) accession self-fertilized for 30 generations in the glasshouse in Jena, Germany and for which extensive knowledge exists regarding the leaf secondary metabolite responses to insect attack. Herbivory by larvae of the specialist lepidopteran *M. sexta* was simulated by applying freshly collected oral secretions to mechanically wounded leaves (35). This procedure, hereafter referred to as W+OS treatment, elicits, in a highly reproducible manner, major changes in the secondary metabolites of *N. attenuata* leaves that can be profiled by UHPLC-ESI/qTOFMS (28, 36). After mass feature (*m/z* signals detected for a specific retention time) extraction, alignment, and retention correction from the overall sample population, the resulting concatenated data matrix (Dataset S1) consisting of 1,044 *m/z* features (not including isotope peak features) was analyzed using principal component analysis (PCA) statistical modeling to explore the variance structure of across-individual metabolic profiles without categorizations by accession location (*SI Appendix*, Fig. S1).

Consistent with the high technical reproducibility of UHPLC-ESI/qTOFMS measurements and of the postprocessing pipeline, U30 samples clustered in close proximity in a central region of the score plot generated from these two principal components (PCs) (*SI Appendix*, Fig. S1). The score plot visualization

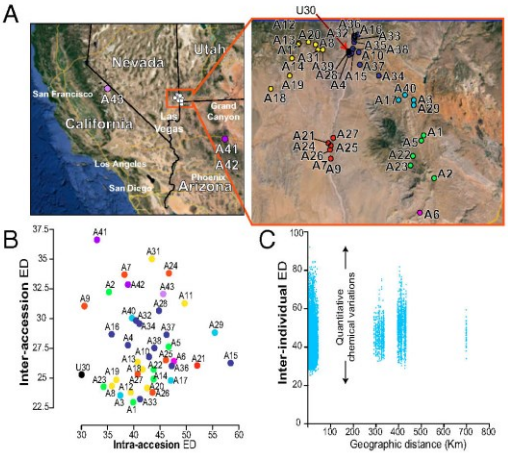

Fig. 1. Herbivory-induced metabolic profiles of *N. attenuata* populations exhibit extensive intra- and interaccession quantitative variations. (*A*) The location of the 43 accession seed collection sites in Utah, Nevada, Arizona, and California. A close-up for the collection sites in Utah is presented. Full GPS coordinates are provided in *SI Appendix*, Table S1. Colors were arbitrarily given to accessions to highlight accessions collected within the same large geographic region. (*B*) Classification of accession according to intra-accession (*x* axis) and interaccession (*y* axis) Euclidean distances calculated for the complete metabolic profile. As expected, metabolomes of replicated plants from U30, a Utah accession inbred for 30 generations, show low variations. Plants from seeds collected in California and Arizona exhibit higher inter- than intraaccession variations. (*C*) Scatter plots of Euclidean distances calculated individual sample pairs, demonstrating no clear relationship between geographical distance and metabolic profile variations. Each dot represents one sample pair, and its coordinates correspond to the geographic distance (*x* axis) and Euclidean distance of the metabolic profiles for this sample pair. Closely related individuals exhibit high quantitative metabolic variation when grown under glasshouse conditions, and no correlation is detected between the geographic and Euclidean distances.

highlighted that there exists a large dynamic range of variation in the relative metabolic composition of the samples analyzed. Consistent with the high variation for within-population genetic structures previously detected via amplified fragment length polymorphism and microsatellite markers (1, 37), we observed that within-accession metabolic samples were in some cases very dissimilar and often did not cluster together within the PCA plot (see PC scores in Dataset S1). We further analyzed sample variations within and among accessions by calculating the respective intra- and interpopulation Euclidean distances (EDs). The ED was used as a consistent estimator for the "metabolic distance" between the complete processed metabolic profiles of a sample pair (Fig. 1*B*). In agreement with the high technical reproducibility of our analytical pipeline, replicate plants from the well-characterized U30 accession exhibited relatively low intraaccession variations ($ED_{intraaccession}$ = 29.3). $ED_{intraaccession}$ values for most other accessions showed much greater variations. This trend was particularly clear for the metabolic profiles of accessions collected in Utah because extremely diverse intraaccession metabolic profiles were detected: $ED_{intraaccession}$ ranging from 29.3 to 58.5. The ED scatter plot presented in Fig.1*B* highlights that, for most accessions, the $ED_{intraaccession}$ was as great as, or even exceeded, variations detected between accessions ($ED_{interaccession}$). We additionally compared the geographic distance between accession sites and metabolic profile divergence calculated for all sample pairs from the dataset ($ED_{interindividual}$). We found no clear relationship between sample pairs' geographic and metabolic distances, and important metabolic variations were observed even within closely collected accessions, as indicated by the large dispersion of $ED_{interindividual}$ values (Fig. 1*C*).

Coexpression Network Analysis Highlights Pathway- and Metabolite-Specific Natural Variation Effects. We next analyzed the relative distribution of metabolite levels within the sample set. To describe in a more systematic manner covariation patterns existing among biochemically related metabolites, we computed a coexpression network [Pearson correlation coefficient (PCC) of >0.75] using the complete dataset of mass features—a mass feature being defined as a deconvoluted *m/z* signal occurring at a given retention time (Fig. 2*A*). Compound class identifiers, W+OS inducibility (fold change W+OS/control in U30, statistic results reported in Dataset S1), and a natural variation estimator [relative median absolute deviation (MAD)] (*Materials and Methods*) were mapped onto this network to annotate regions of the network with low and high degrees of natural variation. Connectivities existing between metabolite-derived signals were then interpreted in light of biochemical relationships and their response to the W+OS induction. Panels of Fig. 2*A* summarize the stepwise process used to pinpointing particular nodes for assessment of their natural variation. The network resolved main natural variation effects within the overall set of detectable mass signals into six main clusters. As discussed in a previous study (38), such groups are formed due to strong and persistent analytical correlations existing between mass signals derived from the ionization and in-source fragmentation of a single metabolite, but also due to the coregulation shared by metabolites of a same metabolic branch. The case of the

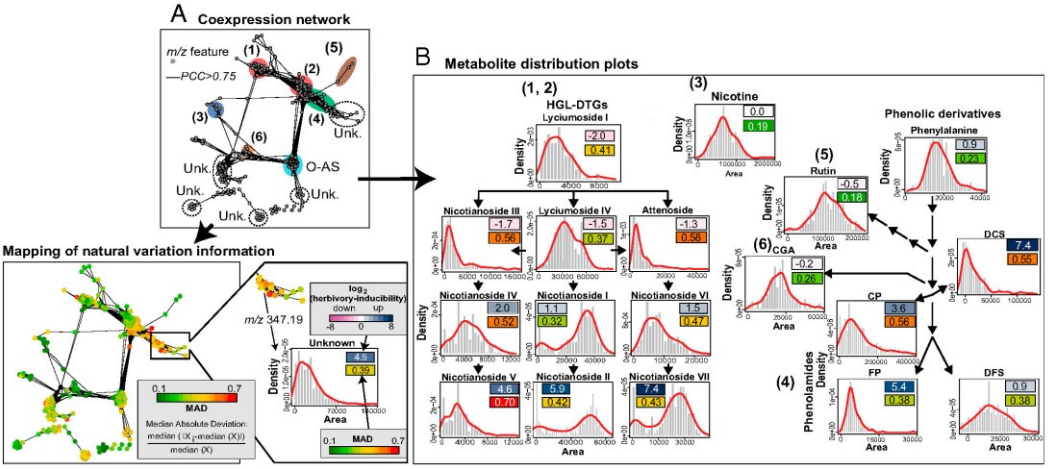

Fig. 2. Natural variation affects metabolite inducibility by herbivory at both pathway and single compound scales. (*A*) Covariance network visualized for metabolite-derived *m/z* features [Pearson correlation coefficient (PCC) of >0.75, biolayout, based on 1,044 *m/z* features extracted from 183 samples] obtained by nontargeted postprocessing of *N. attenuata* accession UHPLC-qTOFMS metabolic profiles indicates pathway-specific and compound-explicit differential inducibility by herbivory and variability through accessions. Different compound classes are annotated using different colored ellipses. Unk., unknown compound class. The resulting network topology was used to map *m/z* feature-specific natural variation coefficients calculated as relative median absolute distance (relative MAD); the degree of MAD is indicated by the node color from green (low) to red (high). In the MAD formula, *Xi* represents the *i*th value across the population for the *m/z* signal denoted as *X*. (*B*) Simplified metabolic schemes and density distribution plots for main intermediates in known secondary metabolic pathways involved in antiherbivore responses in *N. attenuata* that were extracted from the coexpression network: (diagrams 1 and 2) hydroxygeranyllinalool diterpene glycosides (HGL-DTGs); (diagram 3) nicotine; (diagram 4) phenolamides (CP, *N*-caffeoylputrescine; DCS, *N*′,*N*′′-dicaffeoylspermidine; DFS, *N*′,*N*′′-diferuloylspermidine, FP, *N*-feruloylputrescine); (diagram 5) rutin; (diagram 6) CGA, chlorogenic acid. Distribution histogram plots are overlaid with a density line depict distribution of intensities for known metabolites across the dataset with *x* axis, area of intensities, and *y* axis, fitted density with histogram. Nicotine, phenylalanine, rutin, and CGA, which show no to moderate induction by herbivory, exhibit the lowest degree of natural variation in our analysis. Metabolite inducibility is defined as the log_2-scaled fold change value between herbivory-induced and control samples of U30 (average fold change from 13 replicates) and is depicted by color boxes ranging from pink (low) to blue (high). The log_2-scaled inducibility and relative MAD values are reported in the corresponding boxes. Statistics for log_2-scaled fold changes are reported in Dataset S1.

HGL-DTG compounds is particularly germane because these compounds mapped to two locations of the network. This observation may be explained by the fact that HGL-DTG malonylated forms exhibit different temporal dynamics and mode of regulation compared with upstream metabolites (30). Signals corresponding to unknown metabolites were also found adjacent to, and sometimes within, these main network clusters (see example presented for $[M+H]^+$, *m/z* 347.19 in Fig. 2*A*, *Lower*).

To analyze within-pathway effects in more detail, density distribution plots were generated, and MAD scores were used to directly compare scaled distributions across metabolites within defense metabolism pathways (Fig. 2*B*). As previously reported, constitutively produced defense metabolites such as nicotine, rutin, and chlorogenic acid exhibit comparatively low degrees of natural variation. Not only is nicotine a direct defense compound effective against attack by leaf herbivores but its occurrence in the nectar also affects pollinator visits and outcrossing rates (37, 39); as such, stronger selection pressures are expected to erode large quantitative variations in nicotine levels. However, recent work has shown that yet to be identified molecular mechanisms generate high variations in nectar nicotine concentrations that surpass variations found in vegetative tissues (37). This previous result suggests that independent mechanisms may control the variability of nicotine concentrations in these two compartments of a plant. In contrast, high quantitative variations were detected for strongly herbivore-responsive secondary metabolites, especially for metabolites of the phenolamide and HGL-DTG pathways (Fig. 2*B*) as well as *O*-acyl sugar metabolites (*SI Appendix*, Fig. S2). This result is consistent with the fact that metabolic variations detected in this glasshouse-based experiment likely recapitulated plasticity effects of the W+OS treatment on genetically determined metabolic variations. Variations attributable to phenotypic plasticity of different genotypes in response to the W+OS elicitation could in theory result from local adaptations to variations in resources and to the presence or absence of enemies or competitors. Results from Fig. 1 do not fully support this interpretation because W+OS-induced metabolite levels are frequently very variable within an accession. It is noteworthy that the smoke-synchronized germination of seed cohorts of different ages has been shown to result into relatively important within-population genetic diversity levels in this species (24, 25). It is therefore likely that the complex genetic structures frequently detected within small populations also contribute to the maintenance of highly heterogeneous W+OS-induced metabolic responses for a given accession/population.

Interestingly, natural distribution curves for biosynthetically linked metabolites were strikingly malleable, varying from normal distributions to skewed or even bimodal distributions. For simplicity, we describe only a few representative examples. Within the HGL-DTG pathway, variations in lyciumoside I and lyciumoside IV, the two upstream precursors intimately connected with the nonmevalonate primary metabolic pathway (30), were relatively low and were distributed according to a normal distribution-like curve. In clear contrast, gradual increases in distribution diversity were seen for the downstream steps of the pathway. Intermediates in the pathway, such as nicotianoside III and attenoside, distributed along skewed curves and exhibited greater MAD scores than their direct precursors. Highest MAD values were observed for the malonylated HGL-DTGs located most downstream in the pathway, with natural variation patterns being best described by bimodal density distributions. This bimodality is indicative of the polymorphic character of herbivory-induced malonylated HGL-DTG levels in the populations analyzed and is reminiscent of distributions for traits under disruptive selection. The same phenomenon appears when navigating through different ramifications of the phenolic metabolic network with most herbivory-inducible phenolamide derivatives exhibiting high natural variation MAD values whereas rutin and chlorogenic acid (CGA) were visualized as low MAD normal distributions.

Natural Variation in Herbivore-Induced Levels of Known and Unknown Metabolites Partly Overlaps with Variations in Jasmonate Accumulation. The W+OS treatment procedure recapitulates most of the early signaling events activated during *M. sexta* herbivory, including rapid increases in the levels of jasmonates. Jasmonic acid (JA) and its bioactive form jasmonoyl-isoleucine (JA-Ile) regulate most of the changes in metabolism that underlie direct and indirect plant defenses (40–43). Previous work from our group has shown that W+OS-induced levels of these two jasmonates greatly vary in small *N. attenuata* populations growing in the wild and translate into different levels of attractiveness and resistance to natural herbivores (44). As revealed by the shape of the density plots (Fig. 3*A*), JA-Ile displayed greater variations (concentrations ranging from 29.9 to 349.1 ng/g fresh weight) than did JA levels. These large variations may be maintained in natural populations by the counterbalanced effects of JA-Ile–dependent signaling into defense and developmental processes (45). We constructed a correlation map between each of the mass signals and JA and JA-Ile levels for significant PCC values of >0.3 (Fig. 3*B*). As expected, most of the highly inducible secondary metabolites showed strong correlations with either JA or JA-Ile: phenylalanine is highly correlated to JA-Ile but not JA, and HGL-DTGs mainly show high correlations to JA instead of JA-Ile whereas most phenolamides had high PCC scores for both JA and JA-Ile. Interestingly, many unknown metabolites that may participate in the plant defenses also showed significant PCC values with JA or JA-Ile but not with both, indicating that molecular mechanisms underlying their natural variation may deviate from the canonical view of jasmonate signaling: for example, the metabolite with $[M+H]^+$ at *m/z* 350.20 shared a PCC of 0.5 with JA and that at *m/z* 347.19 correlated with JA-Ile with a PCC of 0.38. As previously shown for *N. attenuata* volatile emissions (46), we concluded from this analysis that natural variations in levels of herbivore-induced metabolites only partly overlap with upstream variations in jasmonate accumulation. This result underscores that complex signaling interactions involving not only jasmonates but also other phytohormones and signals vary across accessions (47). A fundamental aspect of this explorative analysis is that it sheds light on many interesting unknown *m/z* signals. These mass signals were part of the network clusters capturing high natural diversity effects and covaried in JA or JA-Ile levels.

Natural Variation Analysis Meets Mass Spectrometry Fragmentation. We designed a workflow to navigate through the *N. attenuata* metabolic space of this experiment (*SI Appendix*, Fig. S3). The workflow is based on shotgun MS/MS data acquisition to collect a holistic repertoire of structural information on the metabolic diversity detected by our analytical platform for this sample set. Shotgun MS/MS indiscriminately considers for fragmentation all signals within an *m/z* range set as large as possible. Data-dependent MS/MS acquisition methods involving the selection of precursor ions for collision-induced dissociation (CID) fragmentation are more frequently applied, but they suffered from several limitations in capturing the metabolic diversity in a natural variation dataset. First, due to scan rate limitation, only a relatively restricted number of precursor ions can be selected for further CID fragmentation in each acquisition cycle, which reduces the comprehensiveness of the MS/MS analysis. Second, the precursor isolation technique is often inaccurate and frequently translates into "contaminated" MS/MS data, leading to low-resolution performance. As proposed by Matsuda et al. (48), this technical limit can be circumvented by using long repetition series for the measurement of one sample and progressively shifting mass ranges selected for CID fragmentation to obtain as

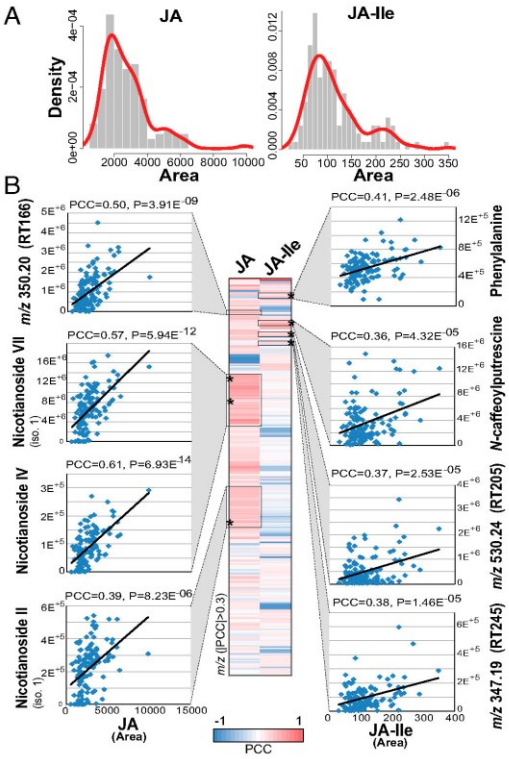

Fig. 3. Natural variation in jasmonate levels only partly accounts for polymorphisms in specialized metabolism and highlights unknown metabolites associated with jasmonate signaling. (*A*) Density distribution plots (*x* axis, area of intensities and *y* axis, fitted density with histogram) (123 samples) illustrating natural variation patterns in JA and JA-Ile levels analyzed by targeted LC-MS/MS/MS for leaf samples collected 1 h after simulated herbivory from glasshouse-grown accessions of *N. attenuata*. (*B*) Heatmap of pairwise Pearson correlation coefficients (PCCs) (only PCCs of >0.3 are shown based on 123 samples) for significant coregulation patterns between deconvoluted *m/z* signals and JA and JA-Ile levels. Examples are presented for known and unknown metabolites with significant correlations with either JA or JA-Ile. Boxes denote in-source fragmentation clusters translating from metabolite ionization and fragmentation. An * indicates the position of metabolite-specific precursor ions from which quantitative data used for the scatter plot representations are derived. iso., isomer; RT, retention time in seconds.

many MS/MS data. However, this approach is extremely time-consuming, and it involves massive data redundancy resulting from iterative acquisition of MS/MS spectra of the same metabolite (20, 49). Shotgun or indiscriminant MS/MS (idMS/MS) analysis, in which every detectable ion is targeted for CID fragmentation, has the main disadvantage of being uninformative about precursor-to-fragment relationships. This method is therefore computationally intensive. However, once these relationships can be assigned confidently within the entirety of a metabolomics dataset, the idMS/MS method is extremely powerful in revealing new structural insights within the entire dataset (20). The concept we propose here of using natural variation patterns for correlation analysis can be extended to all kinds of sample analysis, resulting in sufficient quantitative variations, even for cross-tissue metabolic variation within a single plant. In our workflow (*SI Appendix*, Fig. S3), we selected 10 samples that captured a large proportion of the quantitative and qualitative diversity within the dataset according to the first six components extracted by PCA. These samples were first analyzed under MS1 profile mode and then for idMS/MS with different collision energies. The stepwise assembly of the idMS/MS for rutin, an abundant flavonoid in Solanaceae, is exemplified in *SI Appendix*, Fig. S3. The precursor at *m/z* 611.16 for rutin was correctly annotated by CAMERA (50) as $[M+H]^+$. Briefly, all possible precursor-to-fragment PCC pairs within the retention-dependent compound cluster inferred by deconvolution were separately calculated against the signal at *m/z* 611.16 under four collision energies. Finally, collision energy-specific idMS/MS spectra were merged into a so-called composite idMS/MS spectrum (*SI Appendix*, Fig. S3). This pipeline resulted in 360 deconvoluted idMS/MS spectra (Dataset S2). To summarize, this approach benefits from metabolic natural variation by improving the resolution of PCC analysis and simultaneously capturing rich structural information for metabolites with high variation.

A Biclustering Classification of the idMS/MS Landscape Highlights Structural Features Shared by Metabolic Classes. Most specialized metabolites are taxa-specific and even sometimes species-specific and therefore are not frequently present in public spectral databases that largely include commercially available primary metabolites that are conserved across many organisms. This limitation was clearly verified when querying idMS/MS in Massbank (51) because scores higher than 0.8 were retrieved for only 19% of the idMS/MS (*SI Appendix*, Fig. S4). Recently, the MS/MS molecular network concept has been developed to circumvent the limitation of spectral databases by the analysis of within-dataset MS/MS similarities to formulate a structural hypothesis for unknown MS/MS (22, 52). In the MS/MS molecular network method used for microbial metabolomics (22), network edges represent similarities between MS/MS spectra based on common fragments calculated using a modified normalized dot product (NDP). These previous studies focused on structurally complex microbial specialized metabolites that produced fragment-rich MS/MS spectra that are suitable for fragment-based MS/MS similarity alignment. Here, many of the small molecules we analyzed produced a limited number of fragments, often less than five fragments. A first unsuccessful attempt for idMS/MS classification using only NDP scores based on fragment similarities can be seen in *SI Appendix*, Fig. S5. The most prominent clusters formed in this analysis corresponded to the alignment of the multiple pseudo-MS3 idMS/MS collected for single metabolites being prone to in-source fragmentation during ionization, such as HGL-DTGs. Several recent studies have highlighted the importance of considering neutral losses (NLs) for aligning spectra and constructing similarity clusters overlapping with compound familial groupings. NL analysis has been especially well-implemented in the context of MS/MS fragmentation tree studies (53, 54). We therefore combined these two types of information into a bidimensional clustering method.

The weighted gene coexpression analysis (WGCNA) method is a very powerful method for module clustering based on intercorrelations (55) and has been recently extended to the analysis of differential coexpression with the release of the R package DiffCoEx (56). This approach seemed very efficient for the progressive clustering of our 2D data and for highlighting overlapping and nonoverlapping intercorrelations calculated for either of the two MS/MS similarity measures (Fig. 4*A*). The biclustering by DiffCoEx was conducted on the 360 × 360 NDP similarity (Dataset S2) and 360 × 52 NL similarity score matrices and produced five clustering modules that included 170 idMS/MS spectra. The result of this clustering enables the visualization of relationships between spectrally identical and related metabolites within the measured metabolic landscape (Fig. 4*B*). We color-mapped known and unknown compounds in each of the modules, as well as the intensity of the PCC value with JA and JA-Ile. We

additionally calculated NL overrepresentation for each module by using a chi-square score and depicted NL distribution using a binary heatmap (*SI Appendix*, Fig. S6). Consistently, high NDP and NL scores—overrepresentation of glucose-derived ($P = 1.38 \times 10^{-10}$), rhamnose-derived ($P = 1.07 \times 10^{-33}$), and malonic acid-derived ($P = 4.10 \times 10^{-22}$) NLs—favored the clustering of HGL-DTGs into one module with high PCC for both JA and JA-Ile. Small clusters in this module corresponded to the different idMS/MS or pseudo-MS3 derived from one single ion, which demonstrates that our clustering method allows for the rapid mining of redundancy in MS/MS spectra data collection (*SI Appendix*, Fig. S7). Module 4 that was supported by high NL scores grouped phenolamide, chlorogenic acid, rutin, and nicotine, as well as a large set of unknown metabolites. In the case of phenolamides, NLs were essentially derived from putrescine, spermidine losses (*SI Appendix*, Figs. S5 and S7). Modules 2 and 3 grouped *O*-acyl sugars defined according to different NL compositions. Module 2 shared significant enrichments with module 3 for hexose and methyl pentanoic NLs (*SI Appendix*, Fig. S6). These two modules discriminated type III (module 2) and type II and IV (module 3) of *O*-acyl sugars, which differ by the presence of an acetylated group on the fructose leading to an *m/z* of 204.65 NL, which consequently was detected only in module 2. The first module contained only unknown metabolites that showed great similarities, with both high NDP and NL similarity scores (*SI Appendix*, Fig. S8).

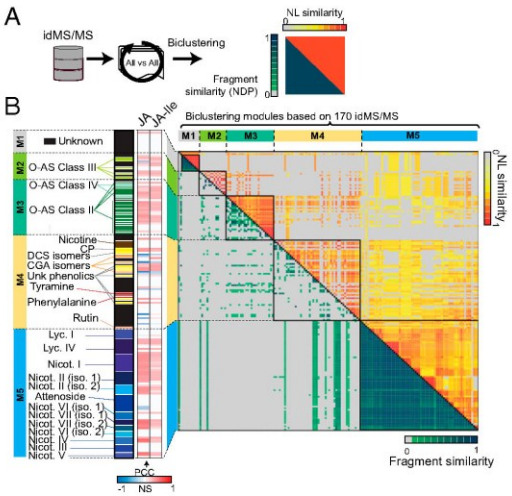

Fig. 4. Biclustering of idMS/MS according to structural relationships computed from fragment and neutral loss similarity metrics facilitate compound class assignments. (*A*) All-against-all alignment of idMS/MS based on fragment and neutral loss similarity calculation. (*B*) Biclustering using the R Diffcoex package of idMS/MS according to results of these two similarity analyses identifies five idMS/MS modules (M1, M2, M3, M4, and M5) that partly overlap to a priori knowledge on compound class definition. Green-to-blue gradient denotes medium-to-high fragment similarity whereas that from yellow-to-red indicates medium-to-high neutral loss similarity. Compound annotation is indicated on the left of the heatmap. Black cells correspond to unknown metabolites whereas the different color variations correspond to different compound classes. The next heatmap bar visualizes significant Pearson correlation values with JA and JA-Ile induced levels as detected in Fig. 3. A neutral loss (NL) map in which shared NLs between classified idMS/MS are reported is presented in *SI Appendix*, Fig. S6. This map was used to infer NLs overrepresented in a particular module. Close views on module subsections highlighting shared NLs and relevant *m/z* features resulting from fragmentation are reported in *SI Appendix*, Figs. S7 and S8. CGA, chlorogenic acid; DCS, *N*′,*N*′′-dicaffeoylspermidine; iso., isomer; Lyc., lyciumoside; Nicot., nicotianoside; O-AS, O-acyl sugars.

Navigating the idMS/MS Molecular Network Facilitates Structural Predictions for Previously Unassigned Herbivory-Regulated Metabolites. We finally illustrate how this clustering can be mined to formulate hypotheses on metabolites with interesting natural variation patterns. Clusters within the biclustering classification heatmap can be selected by making use of the aforementioned mapping of natural variation and jasmonate correlation information and then additionally explored for MS/MS pairwise similarities. This reductionist approach avoids dealing with complex MS/MS molecular networks produced from the overall dataset that are too complex to interpret. Here, module 4 was particularly noteworthy because it was enriched in defensive phenolamides but also included, among others, the unknown compound at *m/z* 347, for which we previously detected an interesting association with jasmonate signaling (Fig. 3*B*) (28, 57). An additional illustration of the mining of module 1 is presented in *SI Appendix*, Fig. S5. From module 4, we constructed a molecular network to formulate hypotheses about this unknown metabolite (Fig. 5*A*) based on the similarity of its idMS/MS with that of known metabolites. For this purpose, we selected all possible NDP and NL similarity-based pairs with a score above 0.6 as network edges and assigned different colors to edges to distinguish NDP and NL connectivities. A subnetwork that contained the idMS/MS for precursor at *m/z* 347.196 ($[M+H]^+$, $C_{19}H_{27}N_2O_4^+$) could then be delimited. The idMS/MS for *m/z* 347 is part of the immediate neighborhood of that of *N*-caffeoylputrescine and of an unknown with $[M+H]^+$ at *m/z* 568.30. This clustering translated from shared NL motifs, such as NLs corresponding to putrescine (*m/z* 88.100) and NH_3 (*m/z* 17.027) and from the presence of a high intensity common fragment corresponding to the cleavage of a caffeic acid moiety from a core molecule (*m/z* 163.039, $C_9H_7O_3^+$, NDP score of 0.32) (Fig. 5 *A* and *B* and *SI Appendix*, Fig. S9). The NL score between these two metabolites was 0.53. Importantly, an idMS/MS corresponding to the CID-induced cleavage of the fragment at *m/z* 259.094 ($C_{15}H_{15}O_4^+$) had also been assembled. idMS/MS for *m/z* 259.094 showed that *m/z* 169.039 and *m/z* 96.055 (C_6H_8O) resulted from the cleavage of this fragment at *m/z* 259.094. This pattern of fragmentation implies that the aromatic ring of the caffeoyl moiety was modified by complexation or acylation with a yet-to-be-determined C_6H_8O residue.

This molecular network-informed analysis is obviously not directly applicable to de novo unknown compound identification, a task that is traditionally achieved after compound purification and de novo identification by NMR. Additionally, this procedure is inherently limited to subsets of the small molecule metabolome because no single analytical procedure can detect the complete metabolome set of a given sample. Nevertheless, such an approach based on molecular networking has been shown to facilitate the process of formulating structural hypotheses by the combined interpretation of phenotypic information and mass spectrometric signatures (22). The "biological contextualization" of these mass spectrometric signatures allows for hypothesis testing using reverse genetics approaches when sufficient knowledge exists about the genes controlling the biosynthesis of a given group of metabolites (Fig. 5*C*). In the context of this study, we first used three transgenic lines produced in the U30 background and in which phenolamide metabolism is affected. When stably silencing *MYB8* (irMYB8), a transcription factor that controls total phenolamide production (28, 29), both *N*-caffeoylputrescine and unknown at *m/z* 347 disappeared from the extract ion current chromatograms (Fig. 5*D*). However, of the two N-acyltransferases (AT1 and DH29) targeted by MYB8, only the silencing by virus-induced gene silencing (VIGs) of *AT1*, which specifically targets phenolic-to-putrescine conjugation, disrupted the production of *N*-caffeoylputrescine and the unknown at *m/z* 347 (*SI Appendix*, Table S3). These

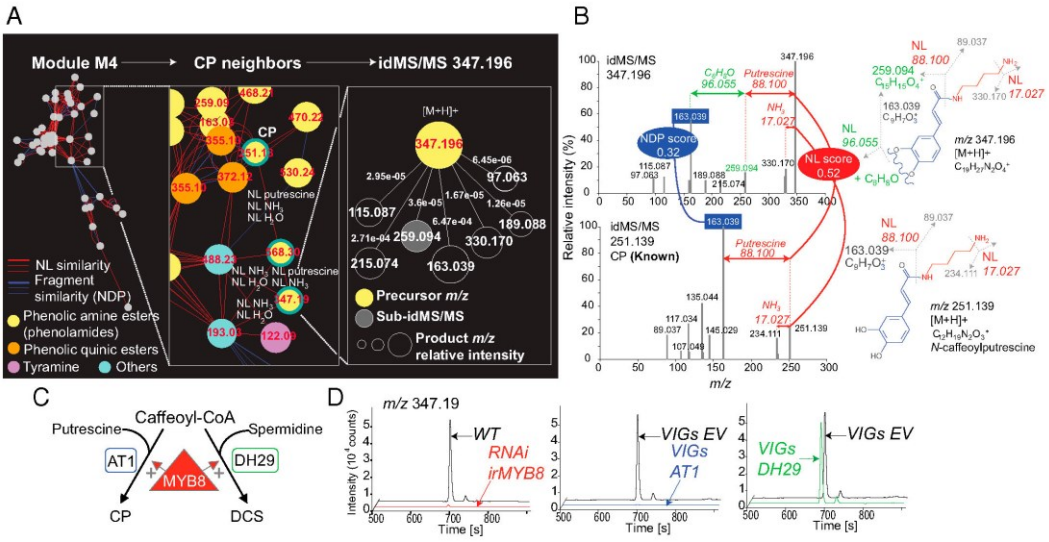

Fig. 5. Navigating the idMS/MS similarity network supports structural predictions for novel herbivory-regulated metabolites. (*A*) Close-up views on the similarity network constructed for module 4 resulting from the biclustering analysis. Module 4 is enriched in previously characterized and structurally elucidated phenolamides, most of which are strongly responsive to simulated herbivory treatments, but also includes unknown metabolites with JA signaling associated natural variation such as *m/z* 347.19 at retention time 245 s. The composite idMS/MS for *m/z* 347.19 is the one of the first network neighbors of *N*-caffeoylputrescine (CP) with idMS/MS *m/z* 251.13 due to neutral loss and fragment-based similarities. (*B*) idMS/MS *m/z* 251.13 and *m/z* 347.19 share neutral loss corresponding to the loss of putrescine. The intense fragment peak at *m/z* 163.04 shared by both idMS/MS corresponds to the caffeoyl moiety cleavage from the putrescine. Interestingly, in the case of idMS/MS *m/z* 347.19, only this fragment derives from an additional neutral loss of C_6H_8O as part of pseudo-MS3 reaction supported by the alignment of idMS/MS *m/z* 259.13. (*C*) Working model for *MYB8*-regulated N-acyltransferase–mediated production of phenolamides. AT1 catalyzes the formation of putrescine-based phenolamides whereas DH29 acts as a first committed step in spermidine conjugate production. (*D*) Extracted ion traces for *m/z* 347.19 supporting its classification as an MYB8-dependent, putrescine-based phenolamide dependent on the catalytic activity of AT1 (*SI Appendix*, Table S3). Additional results obtained from molecular studies of the metabolic conversion from CP to the phenolamide-related *m/z* 347.19 are presented in *SI Appendix*, Fig. S10. EV, empty-vector VIGs control; *irMYB8*, stably silenced *MYB8* transformant; VIGs, virus-induced gene transient silencing.

data, in addition to the strong temporal coregulation upon W+OS treatment of the unknown at *m/z* 347 and *N*-caffeoylputrescine (*SI Appendix*, Fig. S10), provided additional support that this unknown was related to putrescine-based phenolamide metabolism. *N*-caffeoylputrescine is a prominent defense compound in *N. attenuata* (29), and understanding the modulations of its metabolism has important physiological implications for the resistance strategies used by wild populations of this plant. As mentioned above, the fragment at *m/z* 96.055 retrieved in the idMS/MS for *m/z* 347 corresponded to the mass shift and likely the biochemical transformation between *N*-caffeoylputrescine and the unknown at *m/z* 347.196. Because *m/z* 96.055 was predicted as a C_6H_8O residue, we hypothesized that this residue was derived from the fatty acid oxylipin cascade, which converts C18 polyunsaturated fatty acids released from biological membranes during stresses into reactive C6 derivatives. The best-characterized products of this pathway are the green leaf volatiles (GLVs). In *N. attenuata*, stably silencing *LIPOXYGENASE2* (ir*LOX2*), which controls the first committed step in this pathway, fully abolishes C6 aldehydes production and thereby total GLV emissions (58) (*SI Appendix*, Fig. S10). Accumulation of the unknown at *m/z* 347, but not that of *N*-caffeoylputrescine, was almost completely impaired in W+OS-treated ir*LOX2* plants, indicating that LOX2-based fatty acid metabolism was specifically involved in providing the C_6H_8O residue required for the formation of the unknown at *m/z* 347. The exact biochemical reaction involved for the formation of this metabolite is not yet elucidated. Our current hypothesis is that 4-hydroxyhexenal (*SI Appendix*, Fig. S10), one of the most reactive aldehydes produced by this pathway and therefore a pivotal actor in lipid peroxidation-mediated oxidative stress, reacts with *N*-caffeoylputrescine to form the unknown at *m/z* 347. It is unknown yet whether this interaction between C6 metabolism and *N*-caffeoylputrescine benefits the plant by scavenging an excess of highly reactive aldehydes or has homeostatic function over *N*-caffeoylputrescine levels, and/or whether it increases the toxicity of this latter metabolite for herbivores. Additional work, using the panoply of reverse genetic and natural variation resources reported in this study, is needed to investigate chemical and physiological aspects of this intriguing reaction in the context of the plant defense response to highly specialized insect herbivores.

Conclusion

Although heterogeneity in the levels of certain plant metabolites has been analyzed in the context of quantitative genetics approaches to elucidate gene function, little is known about how these variations are organized at the chemical level. There is, therefore, a clear need for workflows that combine biological information and MS data. Ideally, such workflows should comprehensively capture quantitative and structural information on as many detectable metabolites as possible ("metabolic space") and generate data-rich visual outputs to facilitate hypothesis formulation. Here, we implement a workflow fulfilling these requirements and demonstrate that *N. attenuata* populations exhibit large quantitative polymorphisms affecting

secondary metabolism on a single compound but also on a pathway basis. Certain of these variations overlap with variations detected in herbivory-induced levels of JA and JA-Ile.

The coordinated set of information organized by our approach facilitates the mining of known and unknown metabolites, and virtually any kind of biological information (in the present study, natural variation) can be mapped onto MS/MS molecular networks. In this respect, our study complements previous work on molecular networks (22) and exemplifies how these networks can be efficiently mined to formulate structural hypotheses on previously noncharacterized compounds associated with a given phenotype. By combining this structural approach with gene manipulation studies, we notably pinpoint on a biochemical interaction between so-called "direct" (phenolamide metabolism) and "indirect" (C6 metabolism being known to serve for natural enemies' recruitment) metabolism-based defense strategies. The physiological implications of such metabolic cross-talk for a plant's defense response have not yet been explored. To summarize, we predict that the workflow described here will provide some of the essential requirements for a more efficient exploration of the abundant (but not exhaustive) structural information that lies unexplored from most conventional metabolomics screening efforts.

Materials and Methods

Seed Sources. Seeds from *N. attenuata* Torrey ex Watson were collected over the last 20 y by Ian T. Baldwin and his collaborators in the Southwestern United States (*SI Appendix*, Table S1) and were germinated as described in ref. 59. The well-characterized inbred line "UT," which we used as a control comparison, was collected from southwestern Utah in 1996 (population U in ref. 60) and has been self-fertilized for 30 generations in glasshouse conditions in Jena, Germany.

Plant Treatment and Sample Preparation. Metabolic changes induced during *M. sexta* feeding were elicited in a highly synchronized fashion by producing, with a fabric pattern wheel, three rows of punctures onto each side of the midvein of five fully expanded leaves per plant at rosette stage and immediately applying 1:1 diluted *M. sexta* oral secretions (OSs) to the puncture wounds. Four treated leaves were harvested, pooled, and flash-frozen 72 h after elicitation, and metabolites were extracted (*SI Appendix, Materials and Methods*). One treated leaf per plant was harvested 1 h after treatment and analyzed for jasmonates as described in ref. 61.

UHPLC-ESI/TOF-MS Profile Mode Analysis. Two microliters of the 40% methanol leaf extracts were separated using a Dionex rapid separation liquid chromatography system (Dionex) as previously described in ref. 36 (*SI Appendix, Materials and Methods*).

UHPLC-ESI/qTOF-MS Conditions for Indiscriminant MS/MS Data Acquisition. Indiscriminant MS/MS fragmentation analysis (hereafter referred to as idMS/MS) was conducted to gain structural information on the overall metabolic profile detected by UHPLC-TOFMS (*SI Appendix, Materials and Methods*). To increase the chromatographic resolution compared with the conditions previously used for the initial high-throughput analysis of the overall population in the profile mode, we used the following UHPLC binary gradient conditions: 0–1 min, isocratic 90% (vol/vol) A (de-ionized water, 0.1% acetonitrile, and 0.05% formic acid), 10% B (acetonitrile and 0.05% formic acid); 1–22 min, gradient phase to reach 20% A, 80% B; 22–25 min, isocratic 20% A, 80% B. The indiscriminant MS/MS dataset has been deposited in the open metabolomics database Metabolights (www.ebi.ac.uk/metabolights) under accession no. MTBLS203.

Assembly of Compound-Specific idMS/MS. We used the precursor-to-product assignment pipeline developed by ref. 20 and implemented additional rules to improve the accuracy of precursor and product mass signal definition. To reduce the computational and analytical demand, we selected 10 samples from the dataset that capture a great proportion of the variance in the sample population based on manual inspection of chromatograms and on their scores on the six first PCs from the PCA analysis. idMS/MS assembly was achieved via correlational analysis between MS1 and idMS/MS mass signals for low and high collision energy (20) and involved well newly implemented rules (*SI Appendix, Materials and Methods*).

Coexpression Network Construction and Statistical Analysis of Natural Variation Effects. The intradistance for each location was calculated by the average of Euclidean distance crosswise samples within each location, and interdistance for each location was calculated by first computing the average of samples in each location and then computing the average Euclidean distance of all of the other locations calculated with this location in Fig. 1*C*. Euclidean distance in Fig. 1*D* was calculated for each sample pair. The geographic distance was inferred from the Global Positioning System (GPS) coordinates of samples using the great-circle distance algorithm. The degree of variation of each mass feature of the dataset was estimated using the relative median absolute distance calculated as follows:

$$relative\ MAD = \frac{median\ (|Xi - median(X)|)}{median\ (X)}$$

with Xi being ith value across the population for the m/z signal denoted as X.

Coexpression calculations for network construction were computed using the Cytoscape plugin MetaNetter (v2.1) (62, 63).

Pairwise idMS/MS Alignment Based on Fragment Similarity. A standard normalized dot product (NDP), also referred to as cosine correlation method for spectral comparison, was applied using the following equation:

$$NDP = \frac{\left(\sum_i^{S1\&S2} W_{S1,i} W_{S2,i}\right)^2}{\sum_i W_{S1,i}^2 \sum_i W_{S2,i}^2}$$

where $S1$ and $S2$ correspond, respectively, to spectrum 1 and spectrum 2 and $W_{S1,i}$ and $W_{S2,i}$ indicate peak intensity-based weights given to ith common peaks differing by less than 0.01 Da between the two spectra. Weights were calculated as follows:

$$W = [Peak\ intensity]^m [Mass]^n$$

with $m = 0.5$ and $n = 2$, as suggested by MassBank.

Pairwise idMS/MS Alignment Based on Common Neutral Losses. The NL-based similarity between individual idMS/MS was implemented as described in *SI Appendix, Materials and Methods*. We used a list of 52 neutral losses (NLs) commonly encountered during tandem MS fragmentation (*SI Appendix*, Table S2), as well as more specific ones that had been previously annotated for MS/MS spectra of *N. attenuata* secondary metabolite classes.

idMS/MS Molecular Networking by Biclustering. A comparative correlation heat map of 360 idMS/MS spectra was constructed using DiffCoEx (56). The parameters of "cutreeDynamic" were set to cutHeight = 0.999, deepSplit = 1, minClusterSize = 10. The R source code of DiffCoEx was downloaded from additional file 1 in ref. 56, and the required R WGCNA package can be found at labs.genetics.ucla.edu/horvath/htdocs/CoexpressionNetwork/Rpackages/WGCNA/.

ACKNOWLEDGMENTS. We thank Mathias Schöttner for technical support in establishing the idMS/MS acquisition method, Sven Heiling for inspiring discussions, Mario Kallenbach for sharing jasmonate data, and members of the Department of Molecular Ecology for help with sample harvesting. D.L., E.G., and I.T.B. are funded by the Max Planck Society and by Advanced Grant 293926 of the European Research Council (to I.T.B.). E.G.'s research in Heidelberg is supported within the framework of the Deutsche Forschungsgemeinschaft Excellence Initiative to the University of Heidelberg.

1. Xiao Q, et al. (2014) Sources of variability in metabolite measurements from urinary samples. *PLoS ONE* 9(5):e95749.
2. Weng JK (2014) The evolutionary paths towards complexity: A metabolic perspective. *New Phytol* 201(4):1141–1149.
3. Agrawal AA, Hastings AP, Johnson MT, Maron JL, Salminen JP (2012) Insect herbivores drive real-time ecological and evolutionary change in plant populations. *Science* 338(6103):113–116.
4. Berenbaum MR, Zangerl AR (2008) Facing the future of plant-insect interaction research: Le retour à la "raison d'être". *Plant Physiol* 146(3):804–811.
5. Prasad KV, et al. (2012) A gain-of-function polymorphism controlling complex traits and fitness in nature. *Science* 337(6098):1081–1084.
6. Johal GS, Balint-Kurti P, Weil CF (2008) Mining and harnessing natural variation: A little MAGIC. *Crop Sci* 48(6):2066–2073.
7. Toubiana D, et al. (2012) Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations. *PLoS Genet* 8(3):e1002612.
8. Chan EK, Rowe HC, Hansen BG, Kliebenstein DJ (2010) The complex genetic architecture of the metabolome. *PLoS Genet* 6(11):e1001198.

9. Chan EK, Rowe HC, Kliebenstein DJ (2010) Understanding the evolution of defense metabolites in Arabidopsis thaliana using genome-wide association mapping. *Genetics* 185(3):991–1007.
10. Chan EK, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thaliana. *PLoS Biol* 9(8):e1001125.
11. Keurentjes JJ, et al. (2006) The genetics of plant metabolism. *Nat Genet* 38(7): 842–849.
12. Riedelsheimer C, et al. (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc Natl Acad Sci USA* 109(23): 8872–8877.
13. Carreno-Quintero N, Bouwmeester HJ, Keurentjes JJ (2013) Genetic analysis of metabolome-phenotype interactions: From model to crop species. *Trends Genet* 29(1): 41–50.
14. Mitchell-Olds T, Pedersen D (1998) The molecular basis of quantitative genetic variation in central and secondary metabolism in Arabidopsis. *Genetics* 149(2):739–747.
15. Kliebenstein DJ, et al. (2001) Genetic control of natural variation in Arabidopsis glucosinolate accumulation. *Plant Physiol* 126(2):811–825.
16. Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2001) Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in Arabidopsis thaliana leaves and seeds. *Genetics* 159(1):359–370.
17. Keurentjes JJ, Sulpice R (2009) The role of natural variation in dissecting genetic regulation of primary metabolism. *Plant Signal Behav* 4(3):244–246.
18. Keurentjes JJ, et al. (2008) Integrative analyses of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in *Arabidopsis thaliana*. *Genome Biol* 9(8):R129.
19. Carreno-Quintero N, et al. (2012) Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality. *Plant Physiol* 158(3):1306–1318.
20. Broeckling CD, Heuberger AL, Prince JA, Ingelsson E, Prenni JE (2013) Assigning precursor-product ion relationships in indiscriminant MS/MS data from non-targeted metabolite profiling studies. *Metabolomics* 9(1):33–43.
21. Neumann S, Böcker S (2010) Computational mass spectrometry for metabolomics: Identification of metabolites and small molecules. *Anal Bioanal Chem* 398(7-8): 2779–2788.
22. Watrous J, et al. (2012) Mass spectral molecular networking of living microbial colonies. *Proc Natl Acad Sci USA* 109(26):E1743–E1752.
23. Bahulikar RA, Stanculescu D, Preston CA, Baldwin IT (2004) ISSR and AFLP analysis of the temporal and spatial population structure of the post-fire annual, *Nicotiana attenuata*, in SW Utah. *BMC Ecol* 4:12.
24. Baldwin IT, Morse L (1994) Up in smoke: II. Germination of *Nicotiana attenuata* in response to smoke-derived cues and nutrients in burned and unburned soils. *J Chem Ecol* 20(9):2373–2391.
25. Baldwin IT, Staszak-Kozinski L, Davidson R (1994) Up in smoke: I. Smoke-derived germination cues for postfire annual, *Nicotiana attenuata* torr. Ex. Watson. *J Chem Ecol* 20(9):2345–2371.
26. Steppuhn A, Gase K, Krock B, Halitschke R, Baldwin IT (2004) Nicotine's defensive function in nature. *PLoS Biol* 2(8):E217.
27. Steppuhn A, Baldwin IT (2007) Resistance management in a native plant: Nicotine prevents herbivores from compensating for plant protease inhibitors. *Ecol Lett* 10(6): 499–511.
28. Onkokesung N, et al. (2012) MYB8 controls inducible phenolamide levels by activating three novel hydroxycinnamoyl-coenzyme A:polyamine transferases in *Nicotiana attenuata*. *Plant Physiol* 158(1):389–407.
29. Kaur H, Heinzel N, Schöttner M, Baldwin IT, Gális I (2010) R2R3-NaMYB8 regulates the accumulation of phenylpropanoid-polyamine conjugates, which are essential for local and systemic defense against insect herbivores in *Nicotiana attenuata*. *Plant Physiol* 152(3):1731–1747.
30. Heiling S, et al. (2010) Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *Plant Cell* 22(1): 273–292.
31. Macel M, Van Dam NM, Keurentjes JJB (2010) Metabolomics: The chemistry between ecology and genetics. *Mol Ecol Resour* 10(4):583–593.
32. Gulati J, Kim SG, Baldwin IT, Gaquerel E (2013) Deciphering herbivory-induced gene-to-metabolite dynamics in *Nicotiana attenuata* tissues using a multifactorial approach. *Plant Physiol* 162(2):1042–1059.
33. Jansen JJ, et al. (2009) Metabolomic analysis of the interaction between plants and herbivores. *Metabolomics* 5(1):150–161.
34. Marti G, et al. (2013) Metabolomics reveals herbivore-induced metabolites of resistance and susceptibility in maize leaves and roots. *Plant Cell Environ* 36(3):621–639.
35. McCloud ES, Baldwin IT (1997) Herbivory and caterpillar regurgitants amplify the wound-induced increases in jasmonic acid but not nicotine in *Nicotiana sylvestris*. *Planta* 203(4):430–435.
36. Gaquerel E, Heiling S, Schoettner M, Zurek G, Baldwin IT (2010) Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass spectrometry method for induced changes in *Nicotiana attenuata* leaves during simulated herbivory. *J Agric Food Chem* 58(17):9418–9427.
37. Kessler D, et al. (2012) Unpredictability of nectar nicotine promotes outcrossing by hummingbirds in *Nicotiana attenuata*. *Plant J* 71(4):529–538.
38. Gaquerel E, Kuhl C, Neumann S (2013) Computational annotation of plant metabolomics profiles via a novel network-assisted approach. *Metabolomics* 9(4):904–918.
39. Kessler D, Gase K, Baldwin IT (2008) Field experiments with transformed plants reveal the sense of floral scents. *Science* 321(5893):1200–1202.
40. Woldemariam MG, Onkokesung N, Baldwin IT, Galis I (2012) Jasmonoyl-L-isoleucine hydrolase 1 (JIH1) regulates jasmonoyl-L-isoleucine levels and attenuates plant defenses against herbivores. *Plant J* 72(5):758–767.
41. Stitz M, Baldwin IT, Gaquerel E (2011) Diverting the flux of the JA pathway in *Nicotiana attenuata* compromises the plant's defense metabolism and fitness in nature and glasshouse. *PLoS ONE* 6(10):e25925.
42. Kessler A, Baldwin IT (2001) Defensive function of herbivore-induced plant volatile emissions in nature. *Science* 291(5511):2141–2144.
43. Paschold A, Halitschke R, Baldwin IT (2007) Co(i)-ordinating defenses: NaCOI1 mediates herbivore- induced resistance in *Nicotiana attenuata* and reveals the role of herbivore movement in avoiding defenses. *Plant J* 51(1):79–91.
44. Kallenbach M, Bonaventure G, Gilardoni PA, Wissgott A, Baldwin IT (2012) Empoasca leafhoppers attack wild tobacco plants in a jasmonate-dependent manner and identify jasmonate mutants in natural populations. *Proc Natl Acad Sci USA* 109(24): E1548–E1557.
45. Baldwin IT (1998) Jasmonate-induced responses are costly but benefit plants under attack in native populations. *Proc Natl Acad Sci USA* 95(14):8113–8118.
46. Schuman MC, Heinzel N, Gaquerel E, Svatos A, Baldwin IT (2009) Polymorphism in jasmonate signaling partially accounts for the variety of volatiles produced by *Nicotiana attenuata* plants in a native population. *New Phytol* 183(4):1134–1148.
47. Wu J, Hettenhausen C, Schuman MC, Baldwin IT (2008) A comparison of two *Nicotiana attenuata* accessions reveals large differences in signaling induced by oral secretions of the specialist herbivore Manduca sexta. *Plant Physiol* 146(3):927–939.
48. Matsuda F, et al. (2009) MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant J* 57(3):555–577.
49. Hoopmann MR, Finney GL, MacCoss MJ (2007) High-speed data reduction, feature detection, and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass spectrometry. *Anal Chem* 79(15):5620–5632.
50. Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S (2012) CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* 84(1):283–289.
51. Horai H, et al. (2010) MassBank: A public repository for sharing mass spectral data for life sciences. *J Mass Spectrom* 45(7):703–714.
52. Yang JY, et al. (2013) Molecular networking as a dereplication strategy. *J Nat Prod* 76(9):1686–1699.
53. Rasche F, et al. (2012) Identifying the unknowns by aligning fragmentation trees. *Anal Chem* 84(7):3417–3426.
54. Rasche F, Svatos A, Maddula RK, Böttcher C, Böcker S (2011) Computing fragmentation trees from tandem mass spectrometry data. *Anal Chem* 83(4):1243–1251.
55. Langfelder P, Horvath S (2008) WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.
56. Tesson BM, Breitling R, Jansen RC (2010) DiffCoEx: A simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics* 11:497.
57. Gaquerel E, Gulati J, Baldwin IT (2014) Revealing insect herbivory-induced phenolamide metabolism: From single genes to metabolic network plasticity analysis. *Plant J* 79(4):679–692.
58. Allmann S, Halitschke R, Schuurink RC, Baldwin IT (2010) Oxylipin channelling in Nicotiana attenuata: Lipoxygenase 2 supplies substrates for green leaf volatile production. *Plant Cell Environ* 33(12):2028–2040.
59. Krügel T, Lim M, Gase K, Halitschke R, Baldwin IT (2002) Agrobacterium-mediated transformation of *Nicotiana attenuata*, a model ecological expression system. *Chemoecology* 12(4):177–183.
60. Glawe GA, Zavala JA, Kessler A, Van Dam NM, Baldwin IT (2003) Ecological costs and benefits correlated with trypsin protease inhibitor production in *Nicotiana attenuata*. *Ecology* 84(1):79–90.
61. Stitz M, Gase K, Baldwin IT, Gaquerel E (2011) Ectopic expression of AtJMT in *Nicotiana attenuata*: Creating a metabolic sink has tissue-specific consequences for the jasmonate metabolic network and silences downstream gene expression. *Plant Physiol* 157(1):341–354.
62. Jourdan F, Breitling R, Barrett MP, Gilbert D (2008) MetaNetter: Inference and visualization of high-resolution metabolomic networks. *Bioinformatics* 24(1):143–145.
63. Shannon P, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504.

Supporting Information

Navigating natural variation in metabolism: structural analysis of MS/MS profiles from herbivory-induced coyote tobacco populations

Dapeng Li, Ian T. Baldwin, Emmanuel Gaquerel

Materials and Methods

Genetic manipulation of phenolamide-biosynthetic and lipoxygenase genes

The stably-transformed RNAi line silenced for *MYB8* as well virus-induced gene silencing (VIGS) constructs for *AT1* (*N*-acyltransferase producing putrescine-based phenolamides) and DH29 (*N*-acyltransferase producing spermidine-based phenolamides) are described in (1) and were used under the plant growth and inoculation conditions reported in this article. Briefly, 200- to 300-bp fragments of *N. attenuata AT1* and *DH29* genes were amplified by PCR. Amplified fragments were cloned into pTV00 vector, and plasmids were transformed by electroporation into *Agrobacterium tumefaciens* strain GV3101. A pTV00 plasmid without insert (EV) was used as a negative control in all experiments. Three leaves of 24- to 25-d-old *N. attenuata* plants were infiltrated with a 1:1 mixture of *A. tumefaciens* transformed with pBINTRA and one of the pTVAT1, pTVDH29, or pTV00 constructs. Phytoene desaturase (pTVPDS) causing bleaching of tobacco leaves due to the depletion of carotenoids was used as a positive control to monitor the progression of VIGS in a separate set of inoculated plants. After PDS-VIGS leaves developed a strong bleaching phenotype, AT1-, DH29- and EV-VIGS-silenced plants were used for W+OS treatment as for in the case of the natural accession experiment.

Plants stably silenced for LOX2 expression have been notably been characterized by Allmann *et al.* (2). ir*LOX2* do not emit detectable levels of green leaf volatiles. Previous work has shown that this fatty oxidation branch and that of the LOX3-dependent jasmonate biosynthesis function independently and that manipulating one of the two LOX does not significantly alter the

flux through the other LOX-dependent branch (2). The cross between irLOX2 and irLOX3 is deficient in both GLV emissions and jasmonate accumulation.

Metabolite extraction for UHPL-ESI/qTOF-MS analysis

For metabolite profiling, approximately 100 mg of ground leaf tissue was weighed. 1 mL extraction buffer (50 mM acetate buffer, pH 4.8, containing 40 % methanol) per 100 mg tissue was added and samples were homogenized in a ball mill (Genogrinder 2000; SPEX CertiPrep) for 45 s at 1× rate and 250 strokes per min. Homogenized samples were centrifuged at 16,000g, 4°C for 30 min, and supernatants were transferred into 1.5 mL microcentrifuge tubes and re-centrifuged as before. Supernatants of 400 µL were transferred to 2 mL glass vials for mass spectrometry-based metabolomics.

Conditions for UHPLC-ESI/TOF-MS profile mode analysis

The column used was a Acclaim column (150×2.1 mm, particle size 2.2 µm) with a 4 mm×4 mm i.d. guard column of the same material. The following binary gradient was applied using Dionex Ultimate 3000 UHPLC system: 0 to 1 min, isocratic 90% A (de-ionized water, 0.1% [v/v] acetonitrile and 0.05% formic acid), 10% B (acetonitrile and 0.05% formic acid); 1 to 9 min, gradient phase to reach 20% A, 80% B; 9 to 15 min, isocratic 20% A, 80% B. Flow rate was 200 µL/min. Eluted compounds were detected by a high-resolution MicroToF mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an electrospray ionization source operating in positive ionization mode. Typical instrument settings were as follows: capillary voltage 4500 V, capillary exit 130 V, dry gas temperature 200°C, dry gas flow of 8 L/min. Ions were detected from *m/z* 50 to 1400 at a repetition rate of 1 Hz. Mass calibration was performed using sodium formate clusters (10 mM solution of NaOH in 50/50% v/v isopropanol/water containing 0.2% formic acid). Raw data files were converted to netCDF format using the export function of the Data Analysis v4.0 software (Bruker Daltonics, Bremen, Germany).

Additional rules for the assembly of compound-specific idMS/MS

Below are the rules we implemented to improve the correlation calculation output. Rules and the PCC calculation were implemented as a C# script. An issue is that many false positives are produced during the correlation analysis due to the fact that some *m/z* features are only detected in a few samples and that background noise correction can spuriously generate high

correlations. To reduce such false positive errors, we filtered the results of the correlation analysis. To this end, we compared data processing with and without the "fill peaks" function of XCMS (use for background noise correction) and calculated a background noise value from the average correction estimate used by this function to replace "NA" not detected peak intensities. When the "fill peaks" function is used, there still were many "0" intensity values in the dataset which affect the calculation of correlations, and these were replaced with the calculated background value. We also only considered features with intensities that were more than 5 times the background value and considered these as "true peaks". Only *m/z* signals with at least three "true peaks" for the 10 samples precursors (MS1) and fragments datasets were considered for PCC calculation. Logically, *m/z* values for fragments should be lower than that of the precursor and MS/MS fragmentation should occur in the same sample position within the 10 sample dataset as the precursor from which it is derived. Based on these two simple rules, we excluded assigned fragments at *m/z* values larger than that of the identified precursor as well as based the sample position for occurrence precursor and assigned fragments. Finally we merged all the 4 collision energy results for precursor-to-fragment associations into a final deconvoluted spectrum by choosing the highest intensity peak among all candidate peaks of the same *m/z* value at the different collision energies. This is based on the composite spectrum concept and takes into account the different collision energy conditions required to maximize fragmentation possibilities since some fragments can be detected only under certain collision energies. After applying the entire pipeline and set of rules, 360 deconvoluted non-redundant spectra were reconstructed.

Co-expression network construction and statistical analysis of the effects of natural variation

The intra-distance for each location was calculated by the average of Euclidean distance crosswise samples within each location and inter-distance for each location was calculated by first computing the average of samples at each location and then computing the average Euclidean distance of all other locations calculated with this location in **Fig. 1b**. Euclidean distance in **Fig. 1c** was calculated for each sample pair. The geographic distance was inferred from the GPS coordinates of samples using the Great-circle Distance algorithm. The degree of variation of each mass feature of the dataset was estimated using the relative median absolute distance calculated as follows:

$$relative\ MAD = \frac{median\ (|Xi - median(X)|)}{median\ (X)}$$

with *Xi* being *i*th value across the population for the m/z signal denoted as *X*.

Coexpression calculations for network construction were computed using the Cytoscape plugin MetaNetter (v2.1) (3).

idMS/MS molecular networking by biclustering

A comparative correlation heat map of 360 idMS/MS spectra was constructed using DiffCoEx (4). The parameters of "cutreeDynamic" were set to cutHeight = 0.999, deepSplit = 1, minClusterSize = 10. The R source code of DiffCoEx is downloaded from additional file 1 in (4), the required R WGCNA package can be found at: http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/Rpackages/WGCNA.

Tables

Table S1. GPS coordinates of the 43 accessions.

Plant label	Accession	latitude	longitude
1-6	A1	37° 9'28.27"N	113°47'18.16"W
7-11	A2	37° 4'36.82"N	113°46'19.12"W
12-17	A3	37°13'15.83"N	113°48'20.86"W
18-20	A4	37°19'30.66"N	113°57'54.91"W
21-24	A5	37° 9'1.30"N	113°47'43.36"W
25-27	A6	37° 1'2.00"N	113°48'44.10"W
28-29	A7	37° 8'19.58"N	114° 1'35.10"W
30-34	A8	37°20'47.18"N	114° 2'59.84"W
35-36	A9	37° 7'22.15"N	114° 1'36.67"W
37-40	A10	37°19'39.83"N	113°56'27.96"W
41-42	A11	37°20'52.474"N	114° 05'57.799"W
43-47	A12	37°21'02.580"N	114° 05'53.661"W
48-51	A13	37°21'07.103"N	114° 05'50.298"W
52-55	A14	37°19'18.648"N	114°06'53.802"W
56-60	A15	37°19'35.48"N	113°57'38.28"W
61-62	A16	37°21'35.24"N	113°56'38.68"W
63-67	A17	37°13'52.699"N	113° 50'37.113"W
68-72	A18	37°15'28.5"N	114°10'13.3"W
73-77	A19	37°17'09.1"N	114°07'31.5"W
78-82	A20	37°21'12.0"N	114°04'16.7"W
83-88	A21	37°08'35.294"N	114°01'30.659"W
89-91	A22	37°07'03.674"N	113° 49'06.197"W
92-98	A23	37°06'12.5"N	113°49'36.6"W
99-101	A24	37° 8'34.19"N	114° 1'22.95"W
102-106	A25	37° 8'20.56"N	114° 1'29.48"W
107-112	A26	37° 8'16.83"N	114° 1'38.25"W
113-117	A27	37° 9'45.30"N	114° 0'58.52"W
118-120	A28	37°19'33.89"N	113°57'54.16"W
121-122	A29	37°13'5.50"N	113°48'24.25"W
123-135	U30	37°19'36.26"N	113°57'53.05"W
136-138	A31	37°20'22.52"N	114° 2'40.86"W
139-141	A32	37°21'11.21"N	113°57'6.03"W
142-148	A33	37°21'1.04"N	113°57'5.17"W
149-151	A34	37°16'38.65"N	113°53'35.18"W
152	A35	37°20'49.11"N	113°57'7.97"W
153-156	A36	37°21'24.29"N	113°57'0.96"W
157-160	A37	37°18'14.26"N	113°56'23.81"W
161-163	A38	37°19'52.45"N	113°57'31.64"W
164	A39	37°20'9.96"N	114° 2'12.62"W

5

165-168	A40	37°14'27.05"N	113°49'36.71"W
169-170	A41	35°13'7.04"N	111°27'46.10"W
171-172	A42	35°12'56.07"N	111°27'41.29"W
173-183	A43	37°45'19.61"N	118°35'41.82"W

Table S2. List of the 52 neutral losses used for spectral similarity calculations.

The neutral losses used in binary NL matrix generation for DiffCoEx biclustering. In addition to neutral losses that are commonly encountered during tandem MS fragmentation, which corresponded to *N. attenuata* secondary metabolite classes that had been previously annotated for MS/MS spectra, for instance, putrescine and spermidine NL in the case of phenolamides; isobutyric, methyl butanoic and methyl pentanoic acids as well as acetylated fructose losses for *O*-acyl sugars; and finally malonic anhydride and deoxy-hexoses for HGL-DTGs are also included in the NL table. If one of these NLs was detected in an idMS/MS, a score of 1 (otherwise 0) was assigned in order to create a binary vector.

No.	Loss name	Loss formula	Mass difference ±
1	Alkane chains, waxes, fatty acids, methylation	CH_2	14.0157
2	Methane	CH_4	16.0313
3	Ammonium adduct/neutral ammonium loss	NH_3	17.0265
4	Water addition/loss	H_2O	18.0106
5	Salt adduct	K^+ to NH_4^+	20.9293
6	Salt adduct	Na^+ to H^+	21.9819
7	Ethine	C_2H_2	26.0157
8	Carbon monoxide	CO	27.9949
9	Ethene	C_2H_4	28.0313
10	Methylimine	CH_3N	29.0266
11	Formaldehyde	CH_2O	30.0106
12	Methylamine	CH_5N	31.0422
13	Sulfur	S	31.9721
14	Hydrogen sulfide	H_2S	33.9877
15	Salt adduct	K^+ to H^+	37.9559
16	Ketene	C_2H_2O	42.0106
17	Propylation	C_3H_6	42.0470
18	Cyanic acid	CHNO	43.0058
19	Carbon dioxide	CO_2	43.9898
20	Formic acid	CH_2O_2	46.0055

21	Butylation	C_4H_8	56.0626
22	Trimethylamine	C_3H_9N	59.0735
23	Acetic acid	$C_2H_4O_2$	60.0211
24	Urea	CH_4N_2O	60.0324
25	Sulfur dioxide	SO_2	63.9619
26	Pentene	C_5H_8	68.0626
27	Propionic acid	$C_3H_6O_2$	74.0368
28	Benzene	C_6H_6	78.0470
29	Sulfur trioxide	SO_3	79.9568
30	Malonic anhydride	$C_3H_2O_3$	86.0004
31	Isobutyric acid	$C_4H_8O_2$	88.0517
32	Putrescine	$C_4H_{12}N_2$	88.1000
33	Sulfuric acid	H_2SO_4	97.9674
34	Phosphoric acid	H_3PO_4	97.9769
35	Methyl butanoic acid	$C_5H_{10}O_2$	102.0618
36	Malonic acid	$C_3H_4O_4$	104.0110
37	Methyl pentanoic acid	$C_6H_{12}O_2$	116.0861
38	Dihydrogen vinyl phosphate	$C_2H_5O_4P$	123.9926
39	Pentose equivalent	$C_5H_8O_4$	132.0423
40	Spermidine	$C_7H_{19}N_3$	145.1579
41	Deoxy-Hexose equivalent	$C_6H_{10}O_4$	146.0579
42	Hexose-H2O	$C_6H_{10}O_5$	162.0528
43	Deoxy-Hexose equivalent	$C_6H_{12}O_5$	164.0685
44	Glucuronic acid-H2O	$C_6H_8O_6$	176.0321
45	Hexose	$C_6H_{12}O_6$	180.0634
46	Glucuronic acid	$C_6H_{10}O_7$	194.0427
47	Acyl-fructose	$C_8H_{12}O_6$	204.0655
48	Sinapic acid - H2O	$C_{11}H_{10}O_4$	206.0579
49	Glutathione+O-H2O	$C_{10}H_{15}N_3O_6S$	305.0682
50	Glutathione	$C_{10}H_{17}N_3O_6S$	307.0838
51	Sucrose-H2O	$C_{12}H_{20}O_{10}$	324.1057
52	Sucrose	$C_{12}H_{22}O_{11}$	342.1162

Table S3. Comparative analysis of *N*-caffeoylputrescine and unknown at *m/z* 347.19 regulation in AT1 and DH29-VIGs silenced plants.

Data files (4 biological replicates per genotype) from the experiment presented in Fig. 5d were processed by XCMS (cf. script in Dataset S1) to extract peak area values for precursor ions of isomers of *N*-caffeoylputrescine and unknown at *m/z* 347.19. Fold changes between average peak intensity values of empty vector VIGs control samples and of *AT1*- or *DH29*-VIGs silenced samples. Statistics derived from unpaired are additionally reported. Mass features for these metabolites were not detected in ir*MYB8* samples.

m/z	RT (s)	Annot.	*P* value, unpaired t-test		Fold-change		Presence in irMYB8
			AT1-VIGs	DH29-VIGs	AT1-VIGs	DH29-VIGs	
251.134	244	*N*-caffeoyl-putrescine, [M+H]+	0.0000	0.4733	0.16	1.46	**Not detected**
251.134	319	*N*-caffeoyl-putrescine, [M+H]+	0.0000	0.0706	0.18	1.69	**Not detected**
347.193	690	Unidentified *N*-caffeoyl-putrescine metabolite, [M+H]+	0.0000	0.0518	0.20	1.28	**Not detected**
347.193	733	Unidentified *N*-caffeoyl-putrescine metabolite, [M+H]+	0.0000	0.0620	0.14	1.19	**Not detected**

Datasets

Dataset S1. Output of the XCMS and CAMERA processing and statistical analysis and XCMS and CAMERA scripts.

Dataset S2. 360 idMS/MS deconvoluted spectra and NDP and NL matrix used for biclustering. First column corresponds to the 360 idMS/MS precursors in the format of "pcgroup_mz_rt". First row corresponds to the 360 idMS/MS precursors and the 52 NL used for calculation.

Figures

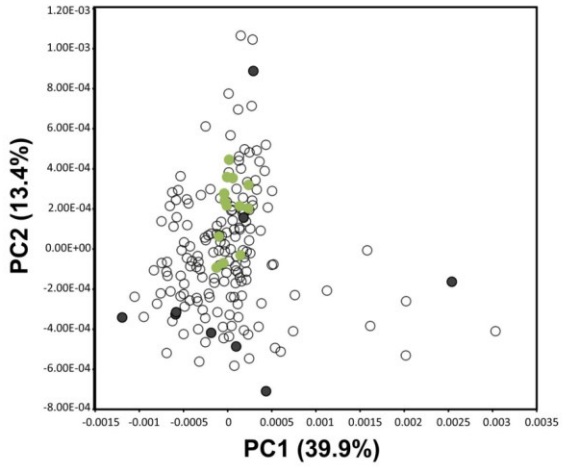

Figure S1. Principal component analysis of the UHPLC-ToFMS metabolic profiles elicited by herbivory for 183 glasshouse-grown plants from 43 natural accessions.

Principal component analysis score plot constructed from autoscaled data derived from the metabolic profiles of 183 herbivory-induced leaf samples of plants grown in the glasshouse from seeds collected in different locations of Southwest of the USA. The two first principal components (PCs) extracted during the modeling accounted for 39.9% and 13.4% of the total variance existing in the sample population. Sample coordinates on the extracted PCs are reported in Dataset S1. Black dots indicate the 10 samples according to PCA-based distant clustering (when considering the first six PCs), geographic origin of the corresponding seeds and manual inspection of chromatographic profiles for metabolic space analysis by indiscriminant MS/MS (see SI Appendix, Fig. S3). Green dots correspond to U30 (Utah accession inbred in the glasshouse for 30 generations) metabolic profiles on the PCA score plot.

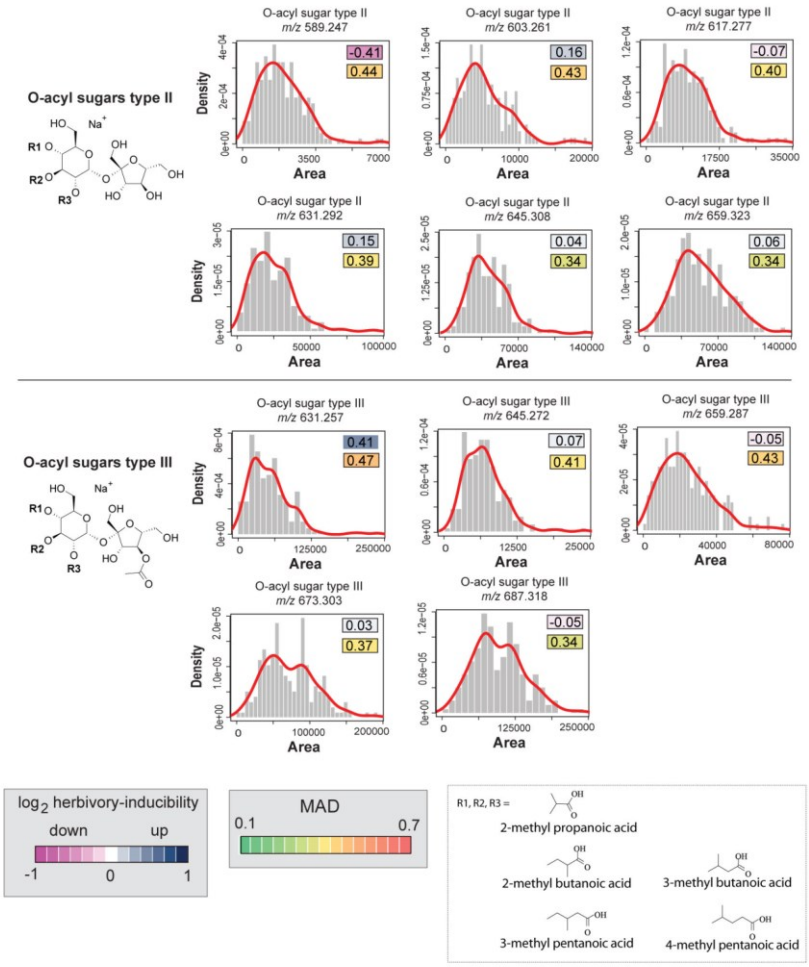

Figure S2. Natural variation analysis for type II and type III *O*-acyl sugars.

Distribution histogram plots are overlaid with a density line depicting distribution of intensities for type II and type III *O*-acyl sugars across 183 individual plants with the x axis corresponding to the area of intensities and the y axis corresponding to the fitted density curve for the histogram. The degree of inducibility for each individual *O*-acyl sugar is denoted by different color from pink (low) to blue (high) as log2-scaled fold change value between herbivory-induced and control samples of U30. The natural variation coefficients are calculated as relative median absolute

distance (relative MAD). The degree of MAD is indicated by the node color from green (low) to red (high). The structures of *O*-acyl sugars are given next to the distribution histogram plots. *O*-acyl sugars are not inducible by the W+OS treatment (5) but natural variations in their abundance were relatively important (MAD ranging from 0.34 to 0.47 for the corresponding $(M+H)^+$ *m/z* signals). Type II and type III *O*-acyl sugars differ in the presence of an acetyl group on their fructose moiety (6, 7, 8). Statistical analysis on fold-change regulation is summarized in Dataset S1.

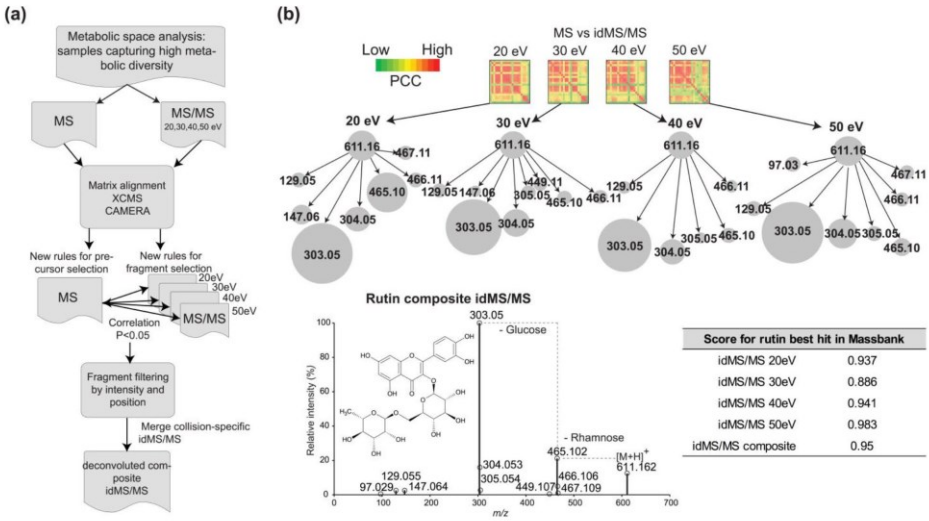

Figure S3. Workflow for metabolic space analysis by shotgun MS/MS and assembly of compound-specific idMS/MS. (a) idMS/MS workfow. To maximize the coverage by shotgun MS/MS of structurally diverse metabolites within a wide dynamic range of concentrations, 10 samples from the clustering analysis of 183 *N. attenuata* accessions were first selected based on the dissimilarity of their processed chemical profiles. These 10 samples were first analyzed by UHPL-qTOFMS using the single MS mode (low fragmentation condition) and then as independent measurements using idMS/MS at 4 different collision energies of 20, 30, 40 and 50ev without any precursor ion selection. Correlation analysis was applied to assign precursor and fragment relationships according to the resulting P-values and newly implemented peak intensity and position filtering functions (see Methods). *m/z* signals co-selected by this pipeline

were further assembled as an individual idMS/MS. **(b)** Correlation-based assembly of rutin (CHEBI:28527) idMS/MS. Heatmaps depict Pearson Correlation Coefficient (PCC) among *m/z* signals for the rutin chromatographic group and the progressive assignment of precursor-fragment relationships. CID voltage-specific (20 to 50ev) precursor-fragment assignments based on PCC are additionally visualized as partial networks with branch length indicative of the PCC significance calculated within the chromatographic group "PCGROUP" suggested by CAMERA and node size is indicative the *m/z* relative intensity. Edge length is inversely proportional to the *P* value: fragments with highly statistically significant precursor-to-fragment PC, the closer the fragments are pulled towards the precursor and the higher the intensity of the fragment, the larger of the space occupied by the circle. Smaller fragments like *m/z* 97.03, are only produced at high collision energy (50ev) and are therefore assigned with a significant correlation. Other fragments are detected only under certain collision energy, for example, *m/z* 449.11 is captured solely for the collision energy of 30ev whereas others, such as *m/z* 465.10, vary in intensity but are ubiquitously detected. To validate the quality idMS/MS assembly, we calculated matching scores with the different collision energy reconstructed spectra against the Massbank database that contains targeted MS/MS measurements on different instruments for rutin standards. All highest scores corresponded to rutin hits and the highest score of 0.983 retrieved from the search was obtained when using the idMS/MS at the 50ev collision energy. The composite idMS/MS for rutin returned a matching score of 0.950 in Massbank.

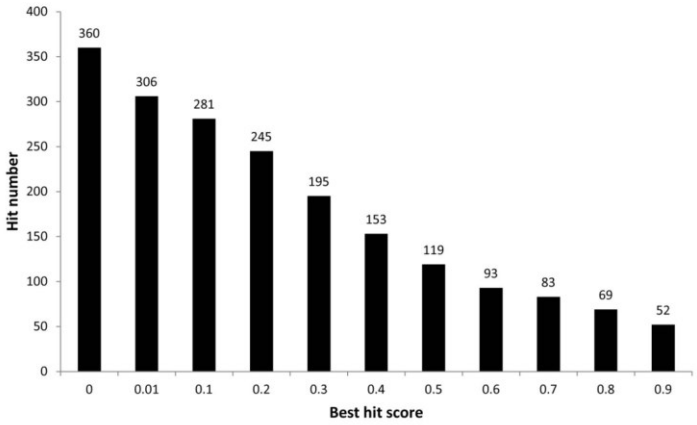

Figure S4. Distribution of idMS/MS score hits during alignment with the MS/MS database Massbank.

Deconvoluted idMS/MS spectra were searched against the MS2 database Massbank. The best hit score from Massbank of each idMS/MS are reported on the x axis. The number of hits that were above of a certain score value is plotted along the y axis.

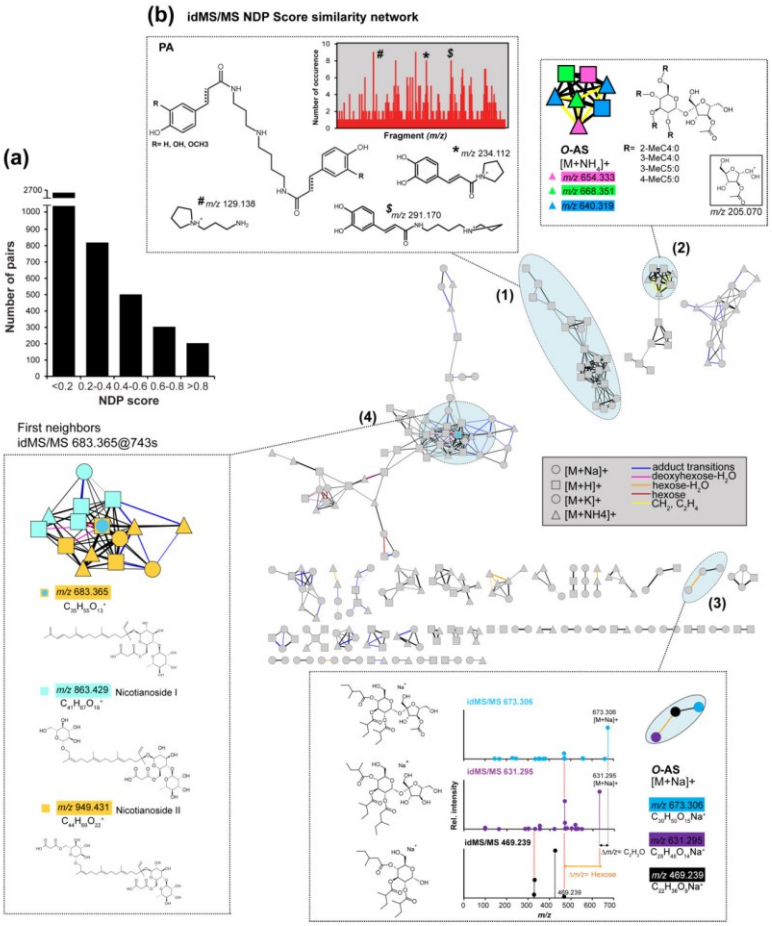

Figure S5. idMS/MS spectral similarity network constructed solely based on pairwise NDP fragment similarity calculations.

(a) Distribution histogram of pairwise NDP scores. **(b)** Molecular network constructed in Cytoscape using the Biolayout for NDP scores > 0.2. A first attempt for high-resolution idMS/MS classification was undertaken using only NDP scores based on fragment similarities. However, since many of the small molecules composing the metabolic profile analyzed produced a limited number of fragments, often < 5 fragments, many pairwise NDP scores are low which does not provide sufficient connectivity for the assembly of a sufficiently resolved molecular

network. The most prominent clusters formed in this analysis corresponded to the alignment of the multiple pseudo-MS3 idMS/MS collected for single metabolites being prone to in-source fragmentation during ionization such as HGL-DTGs but which limits their utility in mining known-to-unknown metabolite relationships. Nevertheless, selected clusters are examined. Cluster **(1)** contains several known phenolamides (PA). This cluster illustrates how groups of known compounds with spectral data classified by NDP fragment similarity can be mined for "overrepresented" m/z fragments. Here, *m/z* signals corresponding to the arrangement of phenolic acyl moieties were more frequently detected than average. As expected, idMS/MS corresponding to the different adduct forms for the same metabolite are also consistently clustered. This is the case of cluster **(2)** which derives from the fragmentation of one *O*-acyl sugar metabolite, in this example the higher than average NDP value results from the shared fragment *m/z* 205.07 (acetylated fructose residue). Cluster **(3)** associates two *O*-acyl sugar metabolites ($[M+H]^+$, *m/z* 673.306 and $[M+H]^+$, *m/z* 631.295). The third idMS/MS (with precursor *m/z* 469.239) corresponds to a pseudo-MS3 of idMS/MS at *m/z* 631.295. Using NDP alone, the clustering of HGL-DTGs is rather difficult to mine. We illustrate for cluster **(4)**, a reconstruction of a directed sub-network for the first neighbors of idMS/MS 683.365, a *m/z* fragment shared by nicotianoside I and II. We conclude that although interesting insights can be visualized from this network, sequences of fragmentation monitored by neutral loss analysis have also to be taken into account when evaluating spectral similarity in order to better assess known-to-unknown metabolite relationships (Fig. 5). Different node symbols denote the idMS/MS precursors for different adduct forms. Edge colors correspond to the most frequently annotated neutral losses.

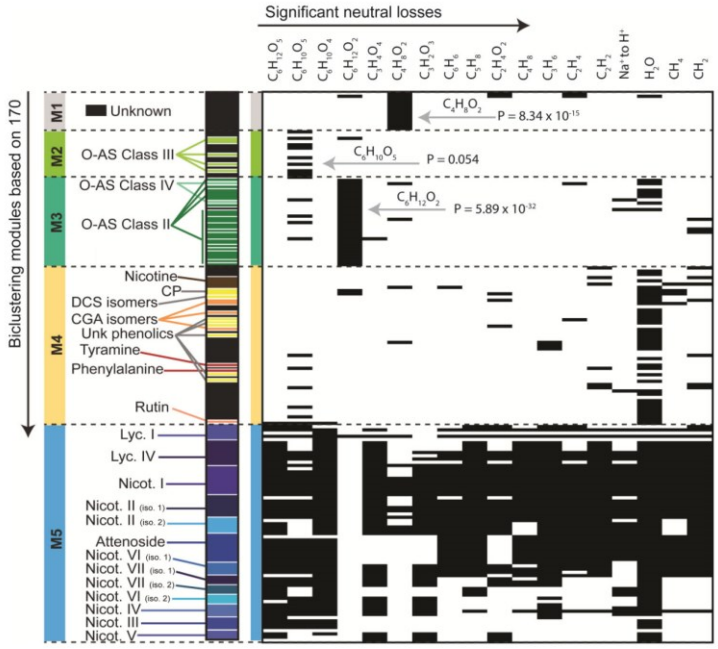

Figure S6. Binary neutral loss map. Black cells within the neutral loss (NL) map (see SI Appendix, Table S2) denote for shared and significantly overrepresented (Chi-Square score analysis) NLs within one module. Statistics for NL significantly overrepresented in modules 1 to 3 are reported for sugar-derived NLs from the backbone structure of *O*-acyl sugars classified into modules 2 and 3. An additional analysis of module 1 is presented in SI Appendix, Fig. S8. As predicted from previous knowledge on their structural characteristics, overrepresentation NLs derived from glucose ($P = 1.38\text{x}10^{-10}$), rhamnose ($P = 1.07\text{x}10^{-33}$) and malonic acid ($P = 4.10\text{x}10^{-22}$)-derived NLs – favored the clustering of HGL-DTGs into one module.

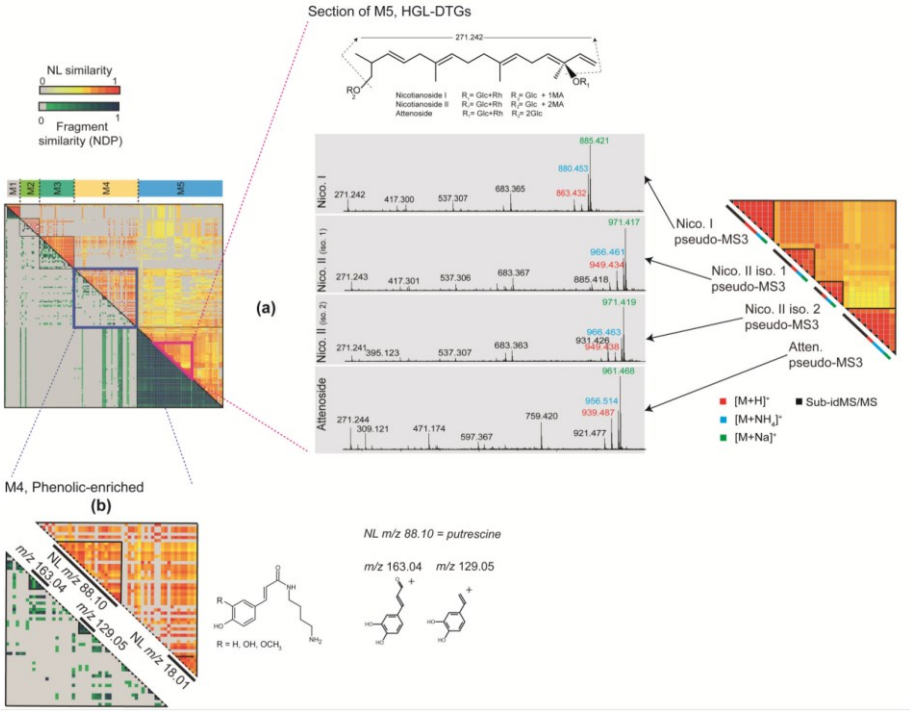

Figure S7. Close-up views on modules 4 and 5 from the biclustering of idMS/MS.

(a) Sub-section of module 5 highlighting the assembly of multiple pseudo-MS3 for different 17-hydroxygeranylinalool diterpene glycosides (HGL-DTGs). HGL-DTG fragmentation has been described by Heiling *et al.* (9). idMS/MS for different adduct forms of single type of HGL-DTGs and the corresponding sub-idMS/MS that derived from in-source fragmentation during ionization are tightly aligned due to high NL and NDP similarities explained by a common backbone-derived fragment at *m/z* 271.242 and common NLs for glucose, rhamnose and malonyl group. **(b)** Close-up view on module 4, previously shown to be enriched in phenolic derivatives. Among others, putrescine-based phenolamide derivatives are clustered due to the presence of a shared NL corresponding to the loss putrescine. High NDP similarities between some of these idMS/MS notably result from a shared fragment corresponding to a caffeoyl moiety. Nico., Nicotianoside; Atten., Attenoside.

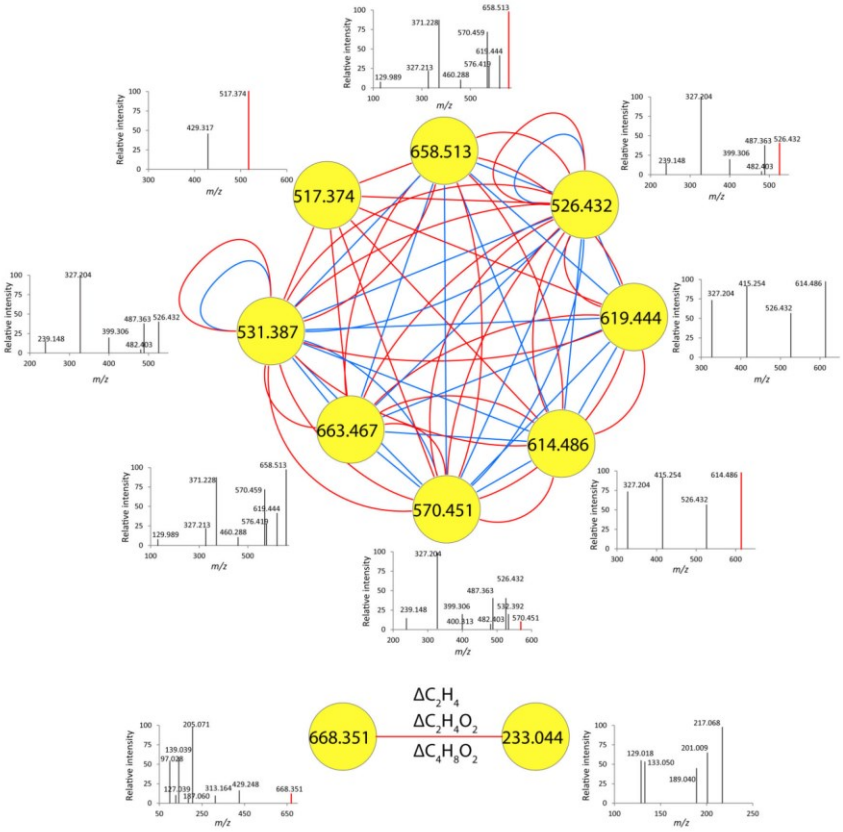

Figure S8. idMS/MS similarity network for unknown metabolites of Module 1 obtained from the biclustering analysis.

Each node represents one precursor of the corresponding idMS/MS spectrum. All possible NDP and NL similarity-based pairs with a score above 0.6 as network edges are assigned with different colors to distinguish NDP and NL connectivities: blue edges correspond to the NDP similarity and red edges to the NL similarity. The approach indicates that Module 1 corresponds likely to three metabolites: $[M+H]^+$, *m/z* 614.486, ($C_{37}H_{64}N_3O_4$); $[M+H]^+$, *m/z* 668.351, ($C_{32}H_{46}N_9O_7$); $[M+H]^+$, *m/z* 233.044, ($C_{12}H_9O_5$). Pseudo-MS3 idMS/MS for the metabolite at $[M+H]^+$, *m/z* 614.486, form a densely connected cluster

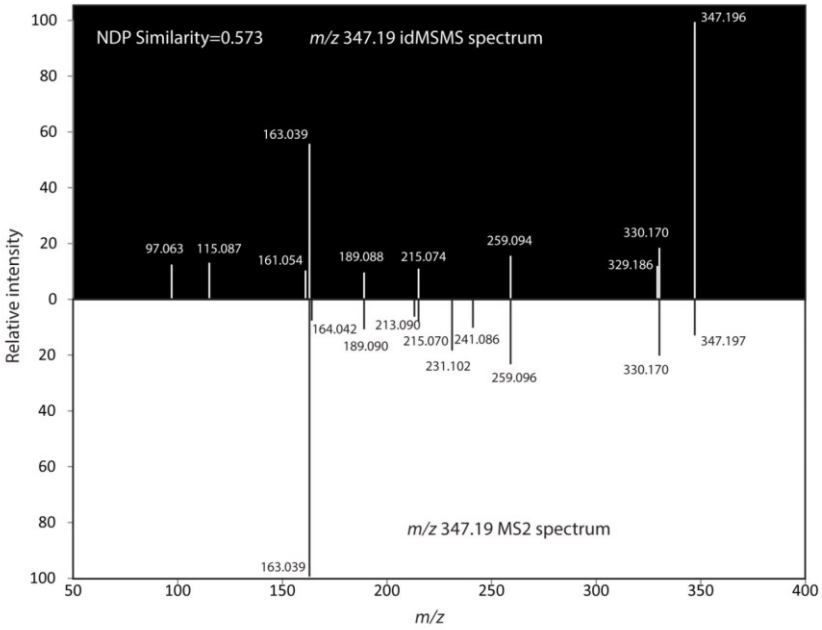

Figure S9. Comparison of the computationally-assembled idMS/MS for *m/z* 347.19 with that obtained by "regular" MS/MS data acquisition.

The experimental idMS/MS spectrum for *m/z* 347.19 is shown (top spectrum) with the corresponding MS2 spectrum obtained with "regular" MS/MS data acquisition using a 0.05 Da mass selection window for a collision energy of 35 eV (bottom spectrum). The similarity score is calculated as the normalized dot product similarity (NDP) score between the experimental idMS/MS spectrum and the MS2 spectrum.

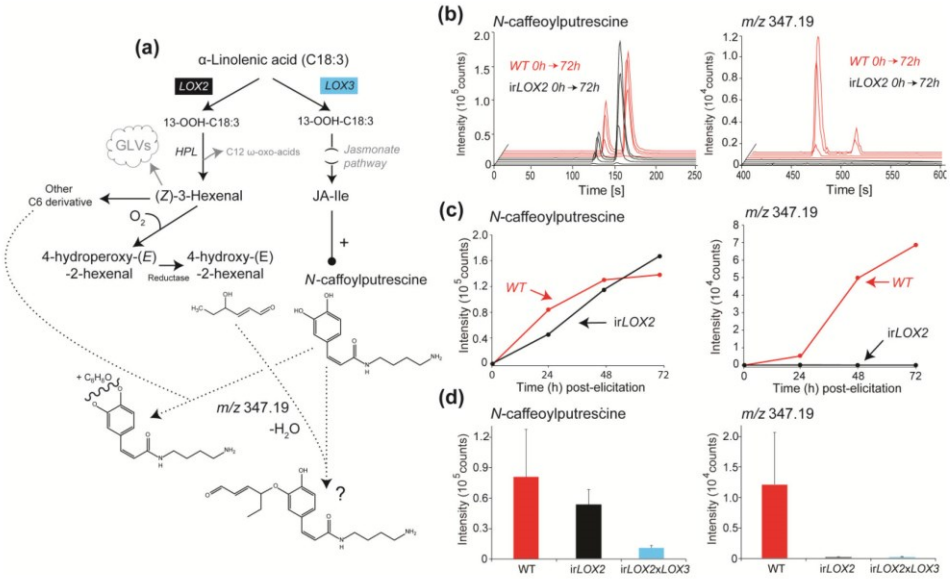

Figure S10. ***LIPOXYGENASE2*-mediated fatty acid oxidation into reactive C6 derivatives is required for the metabolic conversion of *N*-caffeoylputrescine into the unknown at *m/z* 347.19.**

(a) Hypothesis for the interaction between the LOX2-dependent C6 pathway and *N*-caffeoyputrescine metabolism resulting in the formation of the unknown at *m/z* 347.19. Transient *AT1* silencing and the idMS/MS networking approach indicated that the unknown at *m/z* 347.19 is derived from *N*-caffeoylputrescine by a biochemical transformation involving the attachment of a C_6H_8O residue to the caffeoyl moiety. Results in panels **(b-d)** support that this C6 residue derives from the LOX2-dependent fatty acid oxidation pathway. In this pathway, (*Z*)-3-hexenal is produced by HPL-dependent cleavage of linolenic acid 13-hydroperoxides (13-OOH-C18:3) resulting from the activity of LOX2. Previous work has shown that this fatty oxidation branch and that of the LOX3-dependent jasmonate biosynthesis function independently and that manipulating one of the two LOX does not significantly alter the flux through the other LOX-dependent branch (2). (*Z*)-3-hexenal serves for the production of the wide panel of GLVs, but is also abundantly oxidized into 4-hydroperoxy-(*E*)-2-hexenal which is reduced into 4-hydroxy-(*E*)-2-hexenal. This common metabolite, which has been previously detected in *Nicotiana* spp. in both controlled and stress conditions, is highly reactive and toxic for plant cells. Our current

hypothesis is that it reacts (loss of water involved) with *N*-caffeoylputrescine to form the unknown at *m/z* 347. An hypothesis for the product of this reaction is provided, in which 4-hydroxy-(*E*)-2-hexenal reacts with the C3 alcohol. We do not rule out that other LOX2-based C6 derivatives could be involved in the formation of the unknown at *m/z* 347. On the other hand, the LOX3 pathway serves for JA/JA-Ile signaling which regulates at the transcriptional level the biosynthesis of phenolamides including that of *N*-caffeoylputrescine. **(b)** Extracted ion current traces for *N*-caffeoylputrescine isomers and unknown at *m/z* 347.19 accumulation 0, 24, 48 and 72 h after W+OS treatment in WT (red) and ir*LOX2* (black) leaves. **(c)** Mean normalized levels depicting the temporal accumulation *N*-caffeoylputrescine and unknown at *m/z* 347.19 in WT and ir*LOX2*. **(d)** Mean normalized levels *N*-caffeoylputrescine and *m/z* 347.19 in WT, ir*LOX2* and a cross between ir*LOX2* and ir*LOX3* (no C6 and no jasmonate signaling), 72 h after W+OS elicitation. As expected disrupting jasmonate signaling impairs *N*-caffeoylputrescine accumulation and thereby that of unknown at *m/z* 347.19. But, *LOX2* silencing alone is sufficiently to abolish the conversion of *N*-caffeoylputrescine into the unknown *m/z* 347.19 by C_6 residue addition. LOX2, LIPOXYGENASE2; LOX3, LIPOXYGENASE3; HPL, HYDROPEROXIDE LYASE; GLVs, Green Leaf Volatiles.

Supporting References

1. Allmann, S., Halitschke, R., Schuurink, R. C., & Baldwin, I. T. (2010). Oxylipin channelling in *Nicotiana attenuata*: lipoxygenase 2 supplies substrates for green leaf volatile production. *Plant, cell & environment*, 33(12): 2028-2040.
2. Onkokesung N, Gaquerel E, Kotkar H, Kaur K, Baldwin IT, & Galis I. (2012). MYB8 Controls Inducible Phenolamide Levels by Activating Three Novel Hydroxycinnamoyl-Coenzyme A:Polyamine Transferases in *Nicotiana attenuata*. *Plant Physiology* 158: 389-407.
3. Jourdan F, Breitling R, Barrett MP, & Gilbert D (2008) MetaNetter: inference and visualization of high-resolution metabolomic networks. *Bioinformatics* 24(1):143-145.
4. Tesson BM, Breitling R, & Jansen RC (2010) DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC bioinformatics* 11:497.
5. Gaquerel E, Heiling S, Schoettner M, Zurek G, & Baldwin IT (2010) Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass spectrometry method for induced changes in *Nicotiana attenuata* leaves during simulated herbivory. *Journal of agricultural and food chemistry* 58(17):9418-9427.
6. Kim J, et al. (2012) Striking Natural Diversity in Glandular Trichome Acylsugar Composition Is Shaped by Variation at the Acyltransferase2 Locus in the Wild Tomato *Solanum habrochaites*. *Plant physiology* 160(4):1854-1870.
7. Arrendale RF, et al. (1990) Characterization of the Sucrose Ester Fraction from *Nicotiana-Glutinosa*. *Journal of agricultural and food chemistry* 38(1):75-85.
8. Weinhold A & Baldwin IT (2011) Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation. *Proceedings of the National Academy of Sciences of the United States of America* 108(19):7855-7859.
9. Heiling, S., Schuman, M. C., Schoettner, M., Mukerjee, P., Berger, B., Schneider, B., ... & Baldwin, I. T. (2010). Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *The Plant Cell Online*, *22*(1): 273-292.

Manuscript II

Beyond the canon: within-plant and population-level heterogeneity in jasmonate signaling engaged by plant-insect interactions

Dapeng Li, Ian T. Baldwin, and Emmanuel Gaquerel

Published in Plants 2016, 5(1).

plants

MDPI

Review

Beyond the Canon: Within-Plant and Population-Level Heterogeneity in Jasmonate Signaling Engaged by Plant-Insect Interactions

Dapeng Li [1], Ian T. Baldwin [1] and Emmanuel Gaquerel [2,*]

[1] Department of Molecular Ecology, Max Planck Institute for Chemical Ecology, Jena 07745, Germany; dli@ice.mpg.de (D.L.); baldwin@ice.mpg.de (I.T.B.)

[2] Centre for Organismal Studies, University of Heidelberg, Im Neuenheimer Feld 360, Heidelberg 69120, Germany

* Correspondence: emmanuel.gaquerel@cos.uni-heidelberg.de; Tel.: +49-6221-54-5589

Academic Editor: Debora Gasperini

Received: 19 January 2016; Accepted: 9 March 2016; Published: 16 March 2016

Abstract: Plants have evolved sophisticated communication and defense systems with which they interact with insects. Jasmonates are synthesized from the oxylipin pathway and act as pivotal cellular orchestrators of many of the metabolic and physiological processes that mediate these interactions. Many of these jasmonate-dependent responses are tissue-specific and translate from modulations of the canonical jasmonate signaling pathway. Here we provide a short overview of within-plant heterogeneities in jasmonate signaling and dependent responses in the context of plant-insect interactions as illuminated by examples from recent work with the ecological model, *Nicotiana attenuata*. We then discuss means of manipulating jasmonate signaling by creating tissue-specific jasmonate sinks, and the micrografting of different transgenic plants. The metabolic phenotyping of these manipulations provides an integrative understanding of the functional significance of deviations from the canonical model of this hormonal pathway. Additionally, natural variation in jasmonate biosynthesis and signaling both among and within species can explain polymorphisms in resistance to insects in nature. In this respect, insect-guided explorations of population-level variations in jasmonate metabolism have revealed more complexity than previously realized and we discuss how different "omic" techniques can be used to exploit the natural variation that occurs in this important signaling pathway.

Keywords: jasmonate; plant-insect interaction; natural variation; hormone crosstalks; secondary metabolism; metabolomics

1. Introduction

Regulation of plants' defenses during biotic stresses is a complex phenomenon. In nature, plants are essential nutritional sources for different communities of phytophagous insects. The selection pressure imposed by insect herbivory has likely shaped many of the strategies that plants employ to maximize their fitness in their native habitats. In addition to physical barriers (prickles, thorns or sticky trichomes) and constitutive chemical defenses that repel herbivores, plants also rely on more complex defense systems that are specifically produced during insect feeding [1]. These inducible anti-herbivore strategies involving multiple signaling and metabolic layers change a plant's phenotype in order to antagonize insects' growth and/or favor the recruitment of herbivores' natural enemies. Many host plants of chewing insects have evolved the capacity to perceive elicitors in the insect's oral secretions (OS) which triggers signaling pathways and ultimately the synthesis of a repertoire of defensive proteins and secondary metabolites [2–8]. In the case of highly-adapted insects that

evolved counter-adaptations to some of these defensive compounds, the impact of these defenses is frequently limited to the slowing of insect growth or deterrence. As a direct or indirect result, it forces the herbivore to invest in detoxification processes, rather than outright toxicity [9–11]. As a consequence, plants also employ indirect defenses that increase the foraging efficiency of higher trophic levels that prey upon herbivorous insects by, for example releasing herbivory-induced plant volatiles (HIPVs) [9,12–14] or activating tolerance responses involving carbon or nutrient reallocation to less vulnerable plant parts, such as roots [15,16].

Most of the above described anti-herbivore direct and indirect responses are modulated across plant tissues and intrinsically connected with rapid changes in jasmonate pools [17]. Jasmonate biosynthesis and perception have been extensively reviewed [18–24] and are therefore not central to this review. Briefly, jasmonate acid (JA) and its derivatives are synthesized from linolenic acid (18:3) molecules released from chloroplast membranes galactolipids and via a first oxidative reaction catalyzed by lipoxygenase (LOX) proteins. (+)-7-iso-jasmonoyl-L-isoleucine (JA-Ile), the most bioactive jasmonate, has well-established roles in orchestrating many of the physiological processes important to plant-insect interactions [18,25–30]. The perception of jasmonate signaling necessitates the molecular action of the ubiquitin ligase SCF^{COI1} complex via a mechanism in which the F-box protein *COI1 (CORONATINE-INSENSITIVE1)* binds to JA-Ile and recruits *JASMONATE ZIM DOMAIN* proteins (*JAZs*) which are then targeted for ubiquitination and subsequent proteosomal degradation, leading to a release of *JAZ*-mediated repression of gene expression [31–33].

A first dimension of the specialization of the canonical jasmonate signaling lies in the tissue-specific modulation of jasmonates. Other dimensions of variations from the canonical pathway may result from signaling functions of jasmonates other than JA-Ile. 12-oxophytodienoic acid (OPDA) has notably been shown to activate certain physiological responses, such as tendril coiling and carnivorous trap functions and other transcriptional responses in a *COI1*-independent manner [34,35]. The fact that *Arabidopsis* possess 12 *JAZ* genes, some of which are subjected to alternative splicing, provides the basis for diversity in small molecule recognition and signaling that may be responsible for additional heterogeneity in jasmonate signaling amongst tissues [36,37]. An additional dimension in the jasmonate signaling variability space involves the analysis of natural variations existing between geographically distant populations and even in some cases among individuals from populations. In *N. attenuata*, herbivory-induced jasmonate levels greatly vary across geographically-distant accessions and this variation can partly explain differences in levels of secondary metabolite responses [38–41]. In a recent study, we explored the relationships among natural variation in jasmonate signaling and secondary metabolism to gain structural insights into previously unknown jasmonate-dependent secondary metabolites [40]. With the rapid advances in untargeted metabolome analysis, the combination of natural variation with deep metabolomic profiling and compound structural analysis, represent a very promising means of revealing the diversity in jasmonate signaling and the specialized metabolism that it regulates.

In this review, we describe modulations of jasmonate signaling in light of plant-insect interactions as illuminated by examples from recent work with the ecological model, *N. attenuata*. To this end, we provide a three-dimensional roadmap traversing within-plant, population-level to metabolic pathway-specific variations in jasmonate signaling and possible deviations from the canonical perspective. We also discuss means of manipulating jasmonate signaling by creating jasmonate sinks, micrografting procedures and a dexamethasone (DEX) system to disentangle tissue- and age-specificities in jasmonate signaling and their responses. Finally, we discuss the prospect of identifying novel jasmonate-regulated defense metabolites using natural variation in jasmonate signaling. In each of these sections, we highlight how deviations in jasmonate signaling can enhance our understanding of the phenotypic adjustments that are regulated during herbivory by this signaling pathway.

2. Within-Plant Heterogeneity in Jasmonate Signaling and Function

Tissue specificity in metabolism in response to herbivory is thought to be mediated, in part, by tissue-specific deviations in jasmonate signaling [42]. Spatial variations in the expression of the family of *JAZ* genes suggests that the diversity is not simply an artifact of gene redundancy. Despite the current consensus that *JAZ* proteins act as general repressors of jasmonate signaling, little is known about how individual *JAZ* proteins function. Even in the model plant *Arabidopsis*, phenotypes of plants with individually silenced *JAZ* proteins remain incomplete, and their tissue-specific functions remain poorly understood [43,44]. *JAZ* genes are differentially expressed in locally elicited leaves, systemic unelicited leaves and roots of *N. attenuata* when elicited by a standardized method mimicking *Manduca sexta* herbivory (wounding plus application of herbivore oral secretions (W+OS)) [37]. However, this tissue-specific *JAZ* expression is not fully understood and additional work is needed to spatially dissect the function of this *JAZ* regulatory network. Recently, *JAZ10* was successfully employed as a JA-responsive reporter in *Arabidopsis* to identify root-specific activators or repressors of wound signaling [45].

Interactions with other hormonal signaling are also likely to contribute to these tissue-specific responses. For example, specific sets of genes encoding auxin- and ethylene-related signaling processes show OS-elicited responses in systemic leaf tissues that are distinct from those in locally treated tissues [46]. In *Nicotiana tabacum*, jasmonates regulate the expression of *NtPYL4* which encodes an ABA receptor protein affecting the regulation of defensive alkaloids biosynthesis in roots [47]. An antagonistic crosstalk among brassinosteroids (BRs) and jasmonates specifically alters trichome density and defensive metabolite accumulation in tomato [48]. Gibberellins (GAs) interplay with jasmonate signaling through the competitive binding of DELLAs, GA signaling transcriptional repressors, to *JAZ* proteins [49]. The balance of the two negative regulators is flexible according to tissue type [50]. In contrast to the GA-JA synergism in stamen filament development and antagonism in aboveground defense-growth tradeoffs [51,52], a crosstalk between GA and JA has not been yet reported in roots. The interplay between salicylic acid (SA) and jasmonates depends on their tissue-specific concentrations, as SA and JA act synergistically at low concentrations but antagonistically at high concentrations [53]. The kinetics of hormone accumulation after elicitation strongly influences hormonal crosstalk. When SA and JA pathways are activated simultaneously, SA antagonizes JA signaling; however this repression is largely abolished when the two signaling pathways are not synchronously activated [54]. Jasmonate-based defense responses to insect herbivory not only vary across tissues but also throughout the ontogeny of a given tissue as a result of complex rearrangements of these interactions among hormonal pathways [42]. The influence of ontogeny becomes apparent when comparing young with old or even senescing leaves with respect to the amplitude of their herbivory-induced jasmonate bursts and the defense responses they activate. Herbivory-induced jasmonate bursts have been repeatedly shown to be consistently lower in older than in younger tissues. Cytokinins are well known for their roles in inhibiting leaf senescence, a process which jasmonates counter [55,56]; this interplay between the induction of cytokinins and jasmonates in response to herbivory has been shown in both local and systemic tissues [57]. Clearly powerful phenotyping tools combined with tissue- and age-specific genetic manipulations are required to fully address how jasmonate-signaling changes over ontogeny and senescence. A schematic model of jasmonate signaling and responses to insect herbivory in *N. attenuata* is presented in Figure 1.

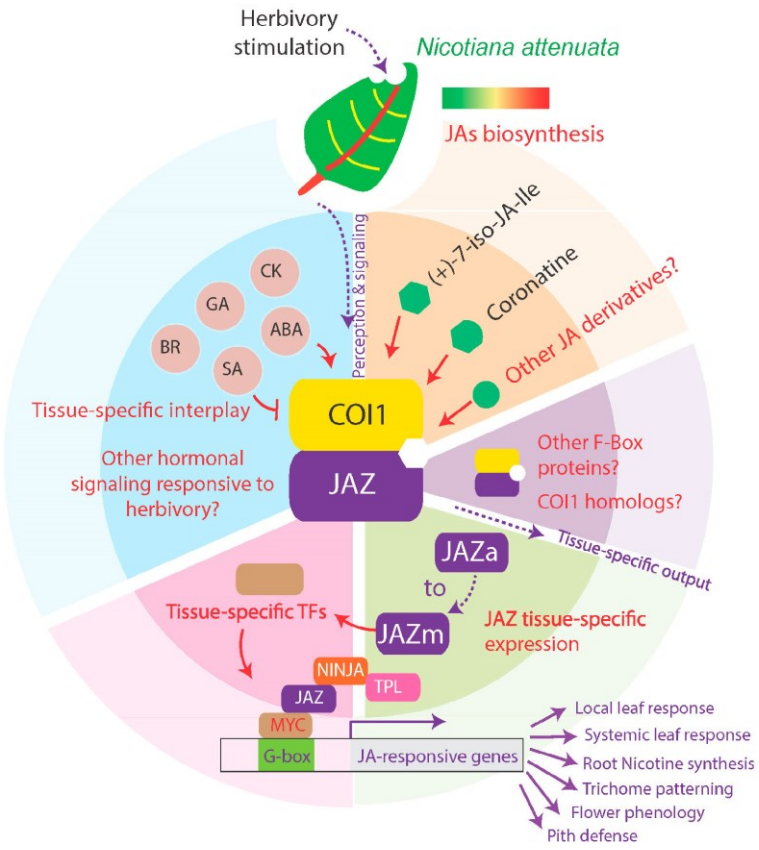

Figure 1. Schematic model of modulations in jasmonate perception, signaling and corresponding tissue-specific herbivory-elicited responses. New discoveries in jasmonate signaling in the ecological model, *Nicotiana attenuata* are presented. The small circle at the top represents jasmonate biosynthesis after herbivore attack; the inner circle represents jasmonate perception and signaling, the outer larger circle represents outputs of tissue-specific responses. The six different types of modulations from the canonical model of jasmonate signaling are presented in different colors. White, jasmonate biosynthesis deviates in different tissue types of leaves; orange, other jasmonate derivatives that are likely to serve as ligands need to be discovered; purple, other F-box proteins may serve as functional groups; green, tissue-specific *JAZ* expressions need to be fully investigated; pink, tissue-specific transcription factors are yet to be discovered; blue, deviations of hormonal crosstalk in different tissues. Red texts indicate areas where modulations are likely to take place. *TPL, TOPLESS.*

2.1. Jasmonate Signaling within-and in Between-Leaves

When chewing herbivores attack a leaf, a short-distance mobile signal is rapidly triggered, which in turn produces a long-distance mobile signal that moves to unattacked leaves following vascular connections. This short-distance mobile signal travels in particular directions towards the undamaged parts of the attacked leaf where it triggers region-specific jasmonate signaling, while the long-distance signal propagates between leaves to elicit jasmonate signaling and defense responses in distal leaves. Relatively little is known about which signal serves as the initial trigger and how the signaling crosses

from wounded leaves to other plant parts. Electric and hydraulic signals are thought to play a central role in this process [58–60]. Recently, a major advance in understanding the mechanisms involved in the electrical component of the signal was achieved when *GLUTAMATE RECEPTOR-LIKE* genes (*GLRs*) were found to be involved in the propagation of electrical signals and to elicit wound-induced expression of several *JAZ* genes [61].

The importance of jasmonate signaling for systemic defensive traits regulation during insect herbivory was first revealed by the seminal work of C.A. Ryan on the induction of digestive proteinase inhibitors in the *Solanum Lycopersicum* tomato [26,27]. Several decades of research from multiple laboratories on the genetic bases of jasmonate metabolism and on the regulation of defense metabolism in insects have firmly established the pivotal role played by jasmonates in orchestrating many of the defenses activated at the whole plant-level that antagonize insect performance [62]. During insect feeding, jasmonate signaling is repeatedly remodeled in both local and systemic tissues by herbivore and damage associated molecular pattern (HAMP and DAMP) elicitors. Jasmonate dynamics detected within- and between-leaf are qualitatively and quantitatively different. In locally-attacked leaves, HAMPs and DAMPs reconfigure the wound response and amplify jasmonate biosynthesis. Local herbivory-induced jasmonate dynamics can be variable across natural populations as well as closely-related species as a result of variations in a plant's ability to perceive HAMP and DAMP elicitors [63,64]. Importantly, rapid changes in jasmonate dynamics occur not only in herbivore-damaged leaves but also in intact systemic tissues [28]. In systemic leaves, jasmonates are synthesized *de novo* in response to transported signal activated by HAMP and DAMP elicitors from damaged leaves [65]. The long distance signaling systems operate at different temporal and spatial scales from those of the short-distance systems. The amount of the jasmonate accumulation and the persistence of the jasmonate levels in undamaged systemic leaves differ from those in wounded leaves [65].

In many species, compositional variations in jasmonate profiles have been observed both within and between leaves. In *Arabidopsis*, jasmonate metabolic profiles differ dramatically both in their compositions and accumulations in systemic leaves and midveins of systemic leaves compared with those in local leaves in response to wounding; while within the wounded leaf, certain jasmonate derivatives specifically accumulate in the midvein, distinguishing the jasmonate signatures from those in the lamina [66]. In *N. attenuata*, the conversion efficiency from JA to JA-Ile greatly differs for different positions relative to the wound site [17]. Constraints in the spread of herbivory-elicited *MAPKs* (*mitogen-activated protein kinases*) activation are responsible for the inhomogeneity of the jasmonate accumulation and transcriptional responses in different areas of the leaves following simulated insect herbivory [67]. Most jasmonate biosynthesis enzymes are present at high levels in vascular bundles [68,69] reinforcing the pivotal roles of midveins and leaf secondary veins in shaping the spatial heterogeneity of jasmonates within a leaf [17]. Along this line, *Arabidopsis LOX6*, a lipoxygenase responsible for rapid jasmonate synthesis in tissues distal to wounds, has been shown to be expressed in a group of cells that are tightly associated with xylem vessels. These regions where contact cells are located are highly sensitive to the release of water column tension in response to wounding. The squeeze cell hypothesis elaborated by E.E. Farmer suggests that contact cells that are hydrostatically coupled to the xylem are likely to be squeezed by pressure increases in the xylem after wounding which disperses hydrostatic pressure finally triggers jasmonate synthesis in distal tissues [70]. Although it is clearly recognized that vascular connectivity among leaves greatly conditions the amplitude of distal jasmonate bursts and dependent-responses, the role of secondary veins in organizing within-leaf spatial spread of jasmonate signaling will requires additional study.

2.2. Root-Mediated Jasmonate Signaling

Root growth inhibition was one of the first observations of a plant's physiological responses to jasmonates and has been used to screen jasmonate-deficiency and insensitivity which led to the discovery of many important signaling components of the jasmonate pathway such as *JAR1*, *COI1*,

JIN4, JIN1/MYC2, JAI3 or *AXR1*. Root-specific molecular aspects of the jasmonate pathway have been recently investigated in depth with the discovery that the organizational principles of the jasmonate signaling machinery differ significantly between the root and the shoot. In this respect, *NINJA* plays a critical role in repressing jasmonate signaling in roots and thereby maintaining the homeostasis of root growth [45]. A follow-up study provided experimental support for a negative regulatory role of *MYC2* together with *NINJA* in constraining jasmonate signaling during normal root cell division and elongation after wounding [71]. These studies provide a molecular framework for understanding the differences in jasmonate signaling between above and belowground organs. In these studies, a *JAZ10* reporter was used to screen for mutants affected in the organ-specific activation of jasmonate signaling in *A. thaliana* seedlings, and highlights the need for tissue-specific manipulations to finely dissect the complexity of jasmonate/*JAZ* regulatory networks.

Relatively little work has been conducted on the role of jasmonate signaling in the regulation of defensive metabolism in roots. The systemic responses in roots during leaf herbivory can exceed the amplitude of some of the best systemic responses detected in aboveground tissues. For instance, the transcriptomic changes activated in roots after *Spodoptera littoralis* herbivory of maize leaves exceeded those observed in damaged leaves [72]. Consistent with this, genes and metabolites that are differentially perturbed between locally damaged and systemically intact tissues in response to *M. sexta* leaf herbivory in *N. attenuata* exhibit more pronounced differential responses in roots than in the shoot, suggesting a root-specific contribution to leaf-based immunity [46]. Jasmonate signaling and dependent responses are not only tissue-specific but also dependent on the systemic organs in which elicited responses take place. The jasmonate metabolic profiles of systemic leaves of root-induced plants varies significantly from the profiles of systemic leaves of shoot-induced plants indicating that roots and shoots likely possess different jasmonate signaling pathways that allow plants to tailor their responses optimally to the organs that are damaged [73,74].

When leaves of tobacco plants are mechanically wounded or attacked by a shoot herbivore, the plants activate synthesis of nicotine in the roots which is then mobilized to the shoot [4]. In *N. attenuata*, a long-distance signal travels from damaged leaves to roots and initiates the transcriptional up-regulation of nicotine biosynthetic genes in the roots [75]. Surprisingly, impairment of jasmonate synthesis and perception in undamaged roots is associated with a priming of jasmonates and abscisic acid (ABA) levels in damaged leaves in response to wounding, indicating that a shoot-root-shoot signaling loop coordinates this whole-plant signaling system [76].This result highlights the important point that to comprehensively elucidate patterns of heterogeneity in jasmonate signaling, an understanding of the integrating signaling networks created by the interactions of multiple hormonal pathways is required.

Many plants exhibit an increase in carbon (C) transport from locally damaged leaves as well undamaged systemic leaves to roots in response to herbivore elicitation [16,77–82]. Such C reallocation might, on the one hand, result in herbivory-induced tolerance via increased C storage and on the other hand, buffer against carbohydrate depletion in roots. Simulated M. sexta herbivory reduces starch and sugar contents in roots of *N. attanuata* which may enable plants to support the synthesis of plant defensive metabolites belowground to attain full defense deployment. Jasmonate signaling plays a key, but not exclusive, function in the regulation of shoot and root carbohydrate turnover.

2.3. Jasmonate Signaling in Tissues Other Than Leaves

Insect herbivory to plant stems results in distinct jasmonate patterns compared to roots and leaves. When *N. attenuata* is attacked by the larvae of the pith-feeding weevil *Trichobaris mucorea*, it activates a separate spatial defensive strategy in which JA and JA-Ile levels specifically increase in the damaged pith region but not in connected leaves (manuscript in preparation) [83]. Another interesting example of tissue-specific modulation of jasmonate signaling during plant-insect interactions concerns the dramatic shifts in flower phenology in *N. attenuata*, a phenomenon which is elicited by *M. sexta* leaf herbivory and requires jasmonate signaling. The larvae of *M. sexta*, the important night pollinator

hawkmoth of *N. attenuata* flowers, are important herbivores and nectaring during pollination increases oviposition rates. This creates a dilemma for *N. attenuata* in how to attract pollinators without increasing herbivory rates. When *N. attenuata* is attacked by *M. sexta* larvae, the phenology of flowers shifts from night-opening and benzyl acetone scenting flowers to morning-opening flowers that do not scent at night and these flowers attract day-active hummingbirds. This herbivory-elicited shift in floral phenology requires jasmonate signaling [84]. Although the underlying regulatory mechanism of the flower opening phenology remains unclear, jasmonates play a central role in floral development and maturation [85], and the speed and rate of corolla opening is highly dependent on jasmonate signaling [86].

Another tissue-specific jasmonate-dependent developmental response related to defense against insects is induced trichome formation. Trichomes often play a role in resistance due to the physical barrier they present to herbivores and for being sites of biosynthesis of various secondary metabolites such as terpenoids, flavonoids and alkaloids that function as direct defenses [87,88] as well as acyl sugars that function as indirect defenses by increasing the predation pressure on attacking herbivores [89]. The tomato mutant *odorless-2* shows altered morphology, density and metabolites in their glandular trichomes, and as a consequence is susceptible to increased attack from both *M. sexta* and Colorado potato beetle larvae in the field [90]. In *Arabidopsis*, endogenous jasmonates regulate trichome initiation via the *COI1* receptor complex to activate the trichome-specific transcription factors *MYB75* and *GLABRA 3* (*GL3*) leading to development of metabolically active trichomes [91,92]. Wound-induced trichome formation is mediated by canonical SCF^{COI1}-dependent jasmonate signaling as *aos* and *coi1*, two mutants that completely lack endogenous jasmonate function show reduced trichome numbers after wounding. Intriguingly, the jar1 mutant which is deficient in the generation of the most bioactive jasmonate compound JA-Ile [93] is unaffected in trichome production. This observation suggests that although JA-Ile is the active ligand of SCF^{COI1}, other active ligands may be responsible for trichome induction [94].

3. Manipulating Tissue-Specific Metabolic Sinks in Jasmonate Signaling

The most commonly used means of studying hormonal function involves, silencing or knockout mutations of key genes in hormonal biosynthesis pathways [95]. Substrate availability also severely determines jasmonate biosynthesis output and creating metabolic sinks has previously been shown to yield clean metabolic and signaling outputs by diverting substrates of the targeted metabolites [96,97]. Manipulating methylation, one of the catalytic reactions used by plants to adjust their pools of hormones, represents a promising approach to unravel jasmonate signaling output and its downstream homeostatic systems [98–100]. Ectopically expressing *AtJMT* gene in *N. attenuata* redirects jasmonate flux towards MeJA which competes with the production of the bioactive form JA-Ile and thereby down-regulates dependent defensive responses [101,102]. This redirection of jasmonate metabolism "silences" jasmonate signaling; the creation of metabolic sinks of jasmonates impairs the accumulation of defense-related transcripts but not transcripts of jasmonate biosynthesis genes and the upstream genes in the jasmonate pathway [101]. With these "jasmonate sink" transformed plants, redirection of jasmonate flux was more pronounced in the midveins and petioles compared with laminas and depleting JA levels clearly accentuated heterogeneities in the jasmonate signatures between local and systemic leaves elicited by simulated herbivory [101]. In *Arabidospsis*, overexpression of *AtJMT* results in elevated levels of MeJA and normal levels of JA but constitutively enhanced expression of JA biosynthesis related genes and plant resistance to stresses [103]. It is noteworthy that *N. attenuata* plants ectopically expressing *AtJMT* do not differ from wild type plants in morphology at the rosette stage of growth. When plants reach flowering stage, the overexpression of *JMT* in *N. attenuata* severely alters floral organ maturation, corolla limb opening and volatile emissions which is consistent with the key function of JA-Ile in regulating these processes [86]. Manipulating tissue-specific jasmonate sinks therefore represents a promising approach to understand the nature of jasmonate homeostasis systems in different organs.

Given the spatiotemporal specificity of jasmonate signaling and the similar specificity in the activity patterns of interacting organisms, it would be ideal to have sufficient control over gene expression to manipulate these dynamic interactions in real time in different tissues and ideally also under field conditions. [104–106] Dexamethasone (DEX) inducible control over gene expression was first applied in *Arabidopsis* to dissect systemic jasmonate signaling [65] and is one of the most sensitive inducible promoter systems available for plant research [107]. Recently, a DEX-inducible pOp6/LhGR expression system was developed for *N. attenuata* and applied to the spatiotemporal modulation of cytokinin signaling during interactions with native herbivore *Tupiocoris notatus* in both laboratory and field studies [65,108]. This DEX-inducible transactivating system contains two transcription units: The first unit employs a constitutive CaMV 35S promoter to express a DEX-responsive chimeric transcription factor (LhGR). The second unit consists of six copies of the transcription factor binding site (lac operators "pop") upstream of a minimal CaMV 35S promotor controlling the expression of inverted repeats of the target gene. In this process, spatial resolution in gene silencing is achieved based on the tissue site of DEX application. As such, this DEX-inducible system allows for surgical manipulations of tissue-specific traits while minimizing off-target effects in other tissues, which may confound interactions with non-targeted organisms [108]. For instance, when inducing the growth of equal-sized lateral branches by apical meristem decapitation, it is possible to achieve branch-specific gene silencing manipulation by the application of the DEX treatment to specific branches without cross-silencing of adjacent branches [108]. This field-applicable DEX-inducible tool when combined with the *JMT/JME* jasmonate sink expression system in *N. attenuata* plants, will provide an ideal means of disentangling tissue-specific regulation of jasmonate signaling for some of the plant-insect interaction-dependent responses described above. A new frontier will consist in the combination of this DEX-transactivating system with tissue- or cell type-specific promoters when relevant ones are identified in *N. attenuata*.

Additionally, (micro)grafting techniques have provided significant insights into the study of systemic signaling. Grafting experiments with tomato mutants in jasmonate biosynthesis and signaling were first employed to reveal that local jasmonate signaling was required to mount systemic responses in distal leaves [109]. Recently, a simple and efficient micrografting method first developed for *Arabidopsis* has been developed to manipulate below- and above-ground parts in *N. attenuata* for the study of shoot-root signaling in plant-insect interactions [110]. The main limitation of this micrografting method is that specific gene silencing is restricted to the root due to the fact that RNAi silencing can move rootward from the shoot. Despite this limitation, the micrografting procedure represents an important advance when combined with silencing/overexpression of genes mediating jasmonate biosynthesis or perception in understanding root- and shoot-specific characterization of gene function [76]. For example, such micrografting combinations of *N. attenuata* lines have been used to reveal that the transportation of nicotine from roots to shoots requires jasmonate synthesis and JA-Ile perception in both shoot and root compartments [76]. Additionally, phytohormone profiling of these micrografting combinations revealed an unexpected elevation of ABA levels in damaged leaves of both micrografted empty vector (EV) (shoot)/RNAi COI1 (root) and EV/RNAi *AOC*. This result highlights the importance of the hormonal crosstalk between JA and ABA in the shoot-root-shoot signaling interplay established during leaf herbivory and induced nicotine mobilization to the shoot.

4. An insect-Guided Tour of Population-Level Polymorphisms in Jasmonate Signaling and Dependent Responses

Considerable amounts of natural variation in jasmonates and dependent responses have been identified in many species [39,41,111,112]. In *Arabidopsis*, MeJA induced gene expressions and glucosinolate accumulation are highly variable among different accessions [111,113]. Modulations of jasmonate accumulation capacities dramatically influence a plant's performance in nature, notably by affecting the inducibility of insecticidal secondary metabolites. For herbivorous insects, the choice of host plants is essential for their survival and reproduction and insects have therefore evolved complex

sensing and behavioral responses to physical and chemical characteristics of their host plants. In this section, we describe interrelationships between modulations of jasmonate signaling and defense metabolism responses and how these affect a plant's interaction with insects as well as insects' host choice. We also discuss how the survey of natural insect populations ("insect-guided tour") can be used to identify natural variations in jasmonate accumulation and signaling [114].

N. attenuata populations exhibit a high genetic diversity within populations, which can be as great as that seen among populations [115] and this is thought to be in part due to the "fire-chasing" germination behavior of this annual plant in which seeds from long-lived seed banks synchronously geminate in response to smoke-derived germination cues [116,117]. This post-fire germination behavior could result in high within-population variation due to the recruitment of plants of very different ages into post-fire populations from the long-lived seed banks. Considerable natural variation in herbivory-induced jasmonate signaling and secondary metabolites have been reported from this species [38–40] and *N. attenuata* genotypes accumulating low amounts of jasmonate are more severely damaged by folivores such as grasshoppers, leafhoppers, lepidopteran larvae, beetles and stem-boring weevils in the field [83,118].

Some phytophagous insects appear to directly sense the jasmonate signaling and/or dependent responses of their host plants to evaluate their suitability as food. For example, the corn earworm *Helicoverpa zea* can respond to jasmonate signals to upregulate the expression of genes important in the detoxification of host plant defenses [119]. *Empoasca* leafhoppers have also evolved mechanisms to select host plants in nature by eavesdropping on their JA-mediated signaling capacities. This probe-feeding insect preferentially targets jasmonate deficient plants [39], but this choice preference is independent of major jasmonate-inducible defense metabolites as well as detectable changes in plant volatiles as revealed from the qualitative and quantitative evaluation of damages from the attack of *Empoasca* leafhoppers on untransformed and jasmonate signaling-deficient transgenic plants under field conditions and the measurement of jasmonate-regulated direct and indirect defense compounds. On the other hand, the special fire-synchronized germination behavior of *N. attenuata* exposes these ephemeral host plants to herbivores from various guilds that colonize the environment after fire, but also to an intense selection for rapid growth with conspecifics and intense competition for resources. Plant's growth-defense trade-offs involving jasmonate regulation might account for the maintenance of low jasmonate-producing genotypes in native populations, providing them with competitive advantages, despite the clear disadvantage of being defenseless and frequently targeted [114].

Jasmonate signaling plays a decisive role in modulating regrowth responses after *M. sexta* elicitation among different natural *N. attenuata* populations and across populations, herbivory induced jasmonate production negatively correlates with flower production of regrowing shoots [120]. Recent studies have revealed additional complexity in the jasmonate signaling pathway and its correlations with the direct and indirect defenses that it regulates. Herbivory-induced concentrations of phytohormones differ significantly among *N. attenuata* accessions, and that individuals even within small populations emit significantly different HIPVs following simulated herbivore elicitation [38] which include trans-α-bergamotene, an abundantly produced herbivory-induced sesquiterpene which functions as an indirect defense in nature. Polymorphisms in endogenous jasmonate signaling among accessions induced by herbivory surprisingly barely correlate with variations in HIPVs profiles. Consistently, when comparing *N. attenuata* Arizona (AZ) and Utah (UT) accessions, the AZ accession was shown to exhibit similar capabilities in producing HIPVs despite having jasmonate accumulations that are half those observed in the UT accession [121,122].

5. Using Natural Variation to Bridge the Gap Between Jasmonate Signaling and Elicited Metabolites

Natural variation has been extensively used to discover the function of genes. With current advances in sequencing and with the use of multi-parent advanced generation inter-cross (MAGIC) and recombinant inbred lines (RILs) combined with analytical approaches such as liquid

chromatography-mass spectrometry (LC-MS) and NMR for targeted metabolic analysis, the exploration of natural variations in both model and crop species can now be used to identify the genetic fundamentals of metabolic traits via quantitative traits locus mapping [123–125]. Large-scale metabolic quantitative trait loci (mQTL) combined with correlation-based network analysis can powerfully query the genomic regions harboring particular secondary metabolic traits [126,127]. Similar approaches can be used in the analysis of large-scale defense metabolism by correlating the variation across different genotypes with the aim of identifying naturally variable and jasmonate-dependent metabolic traits. To implement such an approach, the challenge of the large-scale acquisition of analytical data on as many secondary metabolites as possible for a given tissue needs to be addressed. Hypotheses about the identity of unknown metabolites can be evaluated by patterns of co-regulation for intermediates within a biosynthetic pathway [128]. Such approaches can be progressively expanded to other metabolites beyond those with already known pathway assignments by correlating their variation with those detected at the level of metabolic genes, such as recently implemented in rice and maize and but also for non-model system plants [129–132]. Recently, we developed a metabolomics and computational pipeline using a non-targeted tandem mass spectrometry (MS/MS) approach to resolve the chemodiversity existing in secondary metabolic responses to herbivory elicitation in native populations of *N. attenuata* [40]. The non-targeted MS/MS approach generates non-targeted MS/MS data for all mass signals detected and allows researchers to capture a comprehensive picture of the metabolic diversity detected in these natural populations. Moreover, the use of natural variation patterns across genotypes of a species for a given tissue (in this study, rosette leaves) can support the annotation of previously poorly explored metabolite groups and in theory be extended to other types of sample comparisons, for instance to monitor tissue-wide metabolic variations. Navigating the metabolic natural variation map created in this study showed that both jasmonate metabolism and plants' defensive metabolites elicited by herbivory exhibit high diversity among accessions. However, investigating the correlation between these secondary metabolites induced by herbivory and induced jasmonate levels revealed that the patterns of natural variation detected for these metabolites only partly overlap with upstream variations in jasmonate production [40]. Significant correlations were detected for some of these herbivory-inducible metabolites with either JA or JA-Ile levels but rarely with both. This partial uncoupling between jasmonate variations and quantitative differences in many of the metabolic profiles after herbivory suggests that the jasmonate-mediated metabolic accumulation mechanism is under a complex regulation system which might involve additional components besides the canonical jasmonate signaling pathway (Figure 2).

Non-targeted MS/MS profiling has the potential to identify unknown metabolites that could be involved in anti-herbivore defenses. The problem is that stress-responsive secondary metabolites are extremely structurally diverse and most plant metabolite databases are too sparsely populated to provide useful hints on these unknown structures. Innovative approaches have therefore to be implemented to accelerate the annotation of MS/MS data. One of the successful approaches pioneered by the Dorrestein laboratory involves the creation of MS/MS molecular networks. This approach involves the generation of networks in which mass spectra (nodes) from one data-set are clustered using edges whose length is inversely proportional to the pairwise similarity between spectra. This molecular networking facilitates the mining of unknown metabolites clustered adjacent to known ones. The network view of large metabolomics data-sets also allows for the combined visualization of other biological information linked to a given metabolite (MS/MS spectrum). This possibility was used to visualize statistical descriptors of natural variation for each MS/MS spectrum (nodes in the network) [40], and co-regulation patterns with different jasmonates. More generally, we predicted that such MS/MS network approaches can profoundly advance our understanding of the metabolic heterogeneity associated with modulations of the jasmonate signaling pathway (Figure 2) among different tissue types and/or as result of different stress conditions.

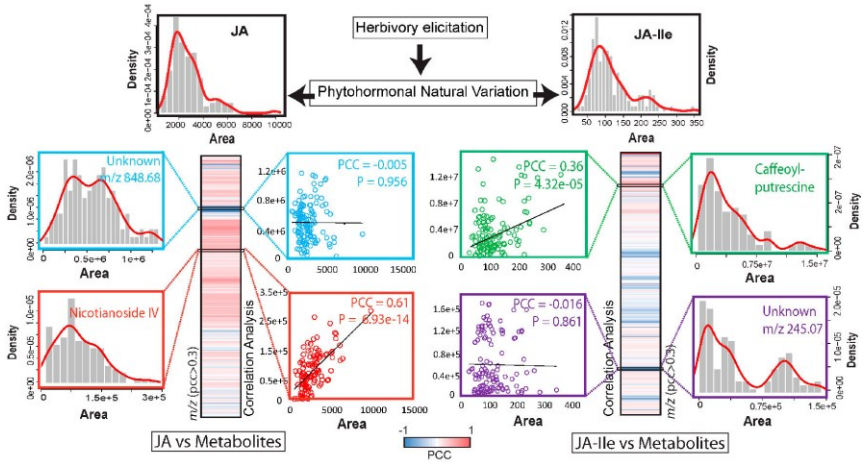

Figure 2. Population-level quantitative variations in herbivore-elicited metabolites only partly overlaps with jasmonate accumulation polymorphisms. The figure is modified from [40]. Density distribution plots of JA and JA-Ile (x axis, area of intensities and y axis, fitted density with histogram) (123 samples) illustrate the patterns of natural variation in JA and JA-Ile levels as analyzed by targeted LC-MS/MS/MS workflows for leaf samples collected 1 h after simulated herbivory from glasshouse-grown accessions of *N. attenuata*. Heatmaps of pairwise Pearson correlation coefficients (PCCs) (only PCCs with either JA or JA-Ile >0.3 are shown) illustrate significant co-regulation patterns between metabolite relative levels and JA and JA-Ile levels. Examples of known and unknown metabolites are depicted in density plots and scatter plots (colored with different color boxes accordingly). The herbivory-inducible defense compound, Nicotianoside IV, correlates significantly with JA whereas N-caffeoylputrescine shows significant correlation with JA-Ile. Unknown *m/z* 848.68 and 245.07 exhibit poor correlations with JA or JA-Ile. Discovering the identity of these and other unknown compounds exhibiting significant correlation scores with JA and JA-Ile levels will be the topic of future research to uncover novel defensive metabolites in *N. attenuata*.

6. Conclusions and Perspectives

More than 50 years of research in jasmonate metabolism and regulation have established the importance of this signaling pathway for a plant's ecological interactions, particularly those with insects. An increasing number of studies demonstrate that tissue-specific responses to herbivory are also associated with tissue-specific regulation in jasmonate signaling. The development of more sensitive and comprehensive analytical tools becomes increasingly important to improve the analysis of temporal modulations across multiple tissues and of the metabolites that are regulated downstream and function as defenses during herbivory. The acquisition of MS/MS metabolite data and the generation of structurally rich MS/MS similarity networks will facilitate the exploration of the multiple metabolic layers that make up plants' defense responses. The analysis of natural variation in metabolic responses harbored in native plant populations will greatly enrich our understanding of the diversity in metabolic outcomes that jasmonate signaling can trigger and provide a healthy antidote to the typological thinking that emerges from focusing too much on models raised in the rarefied environments of growth chambers with little of the environmental complexity that sculpted the diversity of specialized metabolism.

Acknowledgments: Dapeng Li and Ian T. Baldwin are funded by the Max-Planck-Society and by Advanced Grant no. 293926 of the European Research Council to Ian T. Baldwin; Emmanuel Gaquerel's research in Heidelberg is supported within the framework of the Deutsche Forschungsgemeinschaft Excellence Initiative to the University of Heidelberg.

Author Contributions: All authors contributed to the writing, revising and proofreading of the paper.

Conflicts of Interest: The authors declare that there are no conflicts of interest.

References

1. Ballare, C.L. Jasmonate-induced defenses: A tale of intelligence, collaborators and rascals. *Trends Plant Sci.* **2011**, *16*, 249–257. [CrossRef] [PubMed]
2. McConn, M.; Creelman, R.A.; Bell, E.; Mullet, J.E.; Browse, J. Jasmonate is essential for insect defense in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 5473–5477. [CrossRef] [PubMed]
3. Karban, R.; Myers, J.H. Induced plant-responses to herbivory. *Annu. Rev. Ecol. Syst.* **1989**, *20*, 331–348. [CrossRef]
4. Baldwin, I.T. Jasmonate-induced responses are costly but benefit plants under attack in native populations. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 8113–8118. [CrossRef] [PubMed]
5. Moore, B.D.; Andrew, R.L.; Kulheim, C.; Foley, W.J. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytol.* **2014**, *201*, 733–750. [CrossRef] [PubMed]
6. Onkokesung, N.; Gaquerel, E.; Kotkar, H.; Kaur, H.; Baldwin, I.T.; Galis, I. MYB8 controls inducible phenolamide levels by activating three novel hydroxycinnamoyl-coenzyme A: Polyamine transferases in *Nicotiana attenuata*. *Plant Physiol.* **2012**, *158*, 389–407. [CrossRef] [PubMed]
7. Heiling, S.; Schuman, M.C.; Schoettner, M.; Mukerjee, P.; Berger, B.; Schneider, B.; Jassbi, A.R.; Baldwin, I.T. Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *Plant Cell* **2010**, *22*, 273–292. [CrossRef] [PubMed]
8. Halitschke, R.; Schittko, U.; Pohnert, G.; Boland, W.; Baldwin, I.T. Molecular interactions between the specialist herbivore *Manduca sexta* (*Lepidoptera, Sphingidae*) and its natural host *Nicotiana attenuata*. III. Fatty acid-amino acid conjugates in herbivore oral secretions are necessary and sufficient for herbivore-specific plant responses. *Plant Physiol.* **2001**, *125*, 711–717. [PubMed]
9. Schuman, M.C.; Barthel, K.; Baldwin, I.T. Herbivory-induced volatiles function as defenses increasing fitness of the native plant *Nicotiana attenuata* in nature. *Elife* **2012**. [CrossRef] [PubMed]
10. Steppuhn, A.; Gase, K.; Krock, B.; Halitschke, R.; Baldwin, I.T. Nicotine's defensive function in nature. *PLoS Biol.* **2004**, *2*, 1074–1080. [CrossRef] [PubMed]
11. Zavala, J.A.; Giri, A.P.; Jongsma, M.A.; Baldwin, I.T. Digestive duet: Midgut digestive proteinases of *Manduca sexta* ingesting *Nicotiana attenuata* with manipulated trypsin proteinase inhibitor expression. *PLoS ONE* **2008**, *3*, e2008. [CrossRef] [PubMed]
12. Kessler, A.; Baldwin, I.T. Defensive function of herbivore-induced plant volatile emissions in nature. *Science* **2001**, *291*, 2141–2144. [CrossRef] [PubMed]
13. Allmann, S.; Baldwin, I.T. Insects betray themselves in nature to predators by rapid isomerization of green leaf volatiles. *Science* **2010**, *329*, 1075–1078. [CrossRef] [PubMed]
14. Halitschke, R.; Stenberg, J.A.; Kessler, D.; Kessler, A.; Baldwin, I.T. Shared signals—"Alarm calls" from plants increase apparency to herbivores and their enemies in nature. *Ecol. Lett.* **2008**, *11*, 24–34. [CrossRef] [PubMed]
15. Schwachtje, J.; Baldwin, I.T. Why does herbivore attack reconfigure primary metabolism? *Plant Physiol.* **2008**, *146*, 845–851. [CrossRef] [PubMed]
16. Schwachtje, J.; Minchin, P.E.H.; Jahnke, S.; van Dongen, J.T.; Schittko, U.; Baldwin, I.T. SNF1-related kinases allow plants to tolerate herbivory by allocating carbon to roots. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12935–12940. [CrossRef] [PubMed]
17. Stork, W.; Diezel, C.; Halitschke, R.; Galis, I.; Baldwin, I.T. An ecological analysis of the herbivory-elicited JA burst and its metabolism: Plant memory processes and predictions of the moving target model. *PLoS ONE* **2009**, *4*, e4697. [CrossRef] [PubMed]

18. Wasternack, C.; Hause, B. Jasmonates: Biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Ann. Bot.* **2013**, *111*, 1021–1058. [CrossRef] [PubMed]
19. Wasternack, C. How jasmonates earned their laurels: Past and present. *J. Plant Growth Regul.* **2015**, *34*, 761–794. [CrossRef]
20. Wasternack, C. Jasmonates: An update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Ann. Bot.* **2007**, *100*, 681–697. [CrossRef] [PubMed]
21. Kombrink, E. Chemical and genetic exploration of jasmonate biosynthesis and signaling paths. *Planta* **2012**, *236*, 1351–1366. [CrossRef] [PubMed]
22. Shan, X.; Yan, J.; Xie, D. Comparison of phytohormone signaling mechanisms. *Curr. Opin. Plant Biol.* **2012**, *15*, 84–91. [CrossRef] [PubMed]
23. Chini, A.; Boter, M.; Solano, R. Plant oxylipins: COI1/JAZs/MYC2 as the core jasmonic acid-signalling module. *FEBS J.* **2009**, *276*, 4682–4692. [CrossRef] [PubMed]
24. Katsir, L.; Chung, H.S.; Koo, A.J.; Howe, G.A. Jasmonate signaling: A conserved mechanism of hormone sensing. *Curr. Opin. Plant Biol.* **2008**, *11*, 428–435. [CrossRef] [PubMed]
25. Bannenberg, G.; Martinez, M.; Hamberg, M.; Castresana, C. Diversity of the enzymatic activity in the lipoxygenase gene family of *Arabidopsis thaliana*. *Lipids* **2009**, *44*, 85–95. [CrossRef] [PubMed]
26. Farmer, E.E.; Ryan, C.A. Interplant communication—-Airborne methyl jasmonate induces synthesis of proteinase-inhibitors in plant-leaves. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 7713–7716. [CrossRef] [PubMed]
27. Farmer, E.E.; Ryan, C.A. Octadecanoid precursors of jasmonic acid activate the synthesis of wound-inducible proteinase-inhibitors. *Plant Cell* **1992**, *4*, 129–134. [CrossRef] [PubMed]
28. Wu, J.Q.; Baldwin, I.T. New insights into plant responses to the attack from insect herbivores. *Ann. Rev. Genet.* **2010**, *44*, 1–24. [CrossRef] [PubMed]
29. Kazan, K.; Manners, J.M. Jasmonate signaling: Toward an integrated view. *Plant Physiol.* **2008**, *146*, 1459–1468. [CrossRef] [PubMed]
30. Turner, J.G.; Ellis, C.; Devoto, A. The jasmonate signal pathway. *Plant Cell* **2002**, *14*, 153–164.
31. Chini, A.; Fonseca, S.; Fernandez, G.; Adie, B.; Chico, J.M.; Lorenzo, O.; Garcia-Casado, G.; Lopez-Vidriero, I.; Lozano, F.M.; Ponce, M.R.; *et al.* The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **2007**, *448*, 666–664. [CrossRef] [PubMed]
32. Thines, B.; Katsir, L.; Melotto, M.; Niu, Y.; Mandaokar, A.; Liu, G.H.; Nomura, K.; He, S.Y.; Howe, G.A.; Browse, J. JAZ repressor proteins are targets of the SCFCO11 complex during jasmonate signalling. *Nature* **2007**, *448*, 661–662. [CrossRef] [PubMed]
33. Yan, Y.X.; Stolz, S.; Chetelat, A.; Reymond, P.; Pagni, M.; Dubugnon, L.; Farmer, E.E. A downstream mediator in the growth repression limb of the jasmonate pathway. *Plant Cell* **2007**, *19*, 2470–2483. [CrossRef] [PubMed]
34. Ribot, C.; Zimmerli, C.; Farmer, E.E.; Reymond, P.; Poirier, Y. Induction of the *Arabidopsis* PHO1;H10 gene by 12-oxo-phytodienoic acid but not jasmonic acid via a CORONATINE INSENSITIVE1-dependent pathway. *Plant Physiol.* **2008**, *147*, 696–706. [CrossRef] [PubMed]
35. Escalante-Perez, M.; Krol, E.; Stange, A.; Geiger, D.; Al-Rasheid, K.A.S.; Hause, B.; Neher, E.; Hedrich, R. A special pair of phytohormones controls excitability, slow closure, and external stomach formation in the *Venus flytrap*. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15492–15497. [CrossRef] [PubMed]
36. Chung, H.S.; Cooke, T.F.; Depew, C.L.; Patel, L.C.; Ogawa, N.; Kobayashi, Y.; Howe, G.A. Alternative splicing expands the repertoire of dominant JAZ repressors of jasmonate signaling. *Plant J.* **2010**, *63*, 613–622. [CrossRef] [PubMed]
37. Oh, Y.; Baldwin, I.T.; Galis, I. NaJAZh regulates a subset of defense responses against herbivores and spontaneous leaf necrosis in *Nicotiana attenuata* plants. *Plant Physiol.* **2012**, *159*, 769–788. [CrossRef] [PubMed]
38. Schuman, M.C.; Heinzel, N.; Gaquerel, E.; Svatos, A.; Baldwin, I.T. Polymorphism in jasmonate signaling partially accounts for the variety of volatiles produced by *Nicotiana attenuata* plants in a native population. *New Phytol.* **2009**, *183*, 1134–1148. [CrossRef] [PubMed]
39. Kallenbach, M.; Bonaventure, G.; Gilardoni, P.A.; Wissgott, A.; Baldwin, I.T. Empoasca leafhoppers attack wild tobacco plants in a jasmonate-dependent manner and identify jasmonate mutants in natural populations. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1548–1557. [CrossRef] [PubMed]

40. Li, D.P.; Baldwin, I.T.; Gaquerel, E. Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 4147–4155. [CrossRef] [PubMed]
41. Pajerowska-Mukhtar, K.M.; Mukhtar, M.S.; Guex, N.; Halim, V.A.; Rosahl, S.; Somssich, I.E.; Gebhardt, C. Natural variation of potato allene oxide synthase 2 causes differential levels of jasmonates and pathogen resistance in *Arabidopsis*. *Planta* **2008**, *228*, 293–306. [CrossRef] [PubMed]
42. Meldau, S.; Erb, M.; Baldwin, I.T. Defence on demand: Mechanisms behind optimal defence patterns. *Ann. Bot.* **2012**, *110*, 1503–1514. [CrossRef] [PubMed]
43. Pauwels, L.; Goossens, A. The JAZ proteins: A crucial interface in the jasmonate signaling cascade. *Plant Cell* **2011**, *23*, 3089–3100. [CrossRef] [PubMed]
44. Chico, J.M.; Chini, A.; Fonseca, S.; Solano, R. JAZ repressors set the rhythm in jasmonate signaling. *Curr. Opin. Plant Biol.* **2008**, *11*, 486–494. [CrossRef] [PubMed]
45. Acosta, I.F.; Gasperini, D.; Chetelat, A.; Stolz, S.; Santuari, L.; Farmer, E.E. Role of NINJA in root jasmonate signaling. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 15473–15478. [CrossRef] [PubMed]
46. Gulati, J.; Kim, S.G.; Baldwin, I.T.; Gaquerel, E. Deciphering herbivory-induced gene-to-metabolite dynamics in *Nicotiana attenuata* tissues using a multifactorial approach. *Plant Physiol.* **2013**, *162*, 1042–1059. [CrossRef] [PubMed]
47. Lackman, P.; Gonzalez-Guzman, M.; Tilleman, S.; Carqueijeiro, I.; Perez, A.C.; Moses, T.; Seo, M.; Kanno, Y.; Hakkinen, S.T.; Van Montagu, M.C.E.; *et al.* Jasmonate signaling involves the abscisic acid receptor PYL4 to regulate metabolic reprogramming in *Arabidopsis* and tobacco. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5891–5896. [CrossRef] [PubMed]
48. Campos, M.L.; de Almeida, M.; Rossi, M.L.; Martinelli, A.P.; Litholdo, C.G.; Figueira, A.; Rampelotti-Ferreira, F.T.; Vendramim, J.D.; Benedito, V.A.; Peres, L.E.P. Brassinosteroids interact negatively with jasmonates in the formation of anti-herbivory traits in tomato. *J. Exp. Bot.* **2009**, *60*, 4346–4360. [CrossRef] [PubMed]
49. Hou, X.L.; Lee, L.Y.C.; Xia, K.F.; Yen, Y.Y.; Yu, H. DELLAs modulate jasmonate signaling via competitive binding to JAZs. *Dev. Cell* **2010**, *19*, 884–894. [CrossRef] [PubMed]
50. Navarro, L.; Bari, R.; Achard, P.; Lison, P.; Nemri, A.; Harberd, N.P.; Jones, J.D.G. DELLAs control plant immune responses by modulating the balance and salicylic acid signaling. *Curr. Biol.* **2008**, *18*, 650–655. [CrossRef] [PubMed]
51. Cheng, H.; Song, S.S.; Xiao, L.T.; Soo, H.M.; Cheng, Z.W.; Xie, D.X.; Peng, J.R. Gibberellin acts through jasmonate to control the expression of MYB21, MYB24, and MYB57 to promote stamen filament growth in *Arabidopsis*. *PLoS Genet.* **2009**. [CrossRef] [PubMed]
52. Yang, D.L.; Yao, J.; Mei, C.S.; Tong, X.H.; Zeng, L.J.; Li, Q.; Xiao, L.T.; Sun, T.P.; Li, J.G.; Deng, X.W.; *et al.* Plant hormone jasmonate prioritizes defense over growth by interfering with gibberellin signaling cascade. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1192–1200. [CrossRef] [PubMed]
53. Mur, L.A.J.; Kenton, P.; Atzorn, R.; Miersch, O.; Wasternack, C. The outcomes of concentration-specific interactions between salicylate and jasmonate signaling include synergy, antagonism, and oxidative stress leading to cell death. *Plant Physiol.* **2006**, *140*, 249–262. [CrossRef] [PubMed]
54. Thaler, J.S.; Fidantsef, A.L.; Bostock, R.M. Antagonism between jasmonate- and salicylate-mediated induced plant resistance: Effects of concentration and timing of elicitation on defense-related proteins, herbivore, and pathogen performance in tomato. *J. Chem. Ecol.* **2002**, *28*, 1131–1159. [CrossRef] [PubMed]
55. Richmond, A.E.; Lang, A. Effect of kinetin on protein content and survival of detached Xanthium leaves. *Science* **1957**, *125*, 650–651. [CrossRef]
56. He, Y.H.; Fukushige, H.; Hildebrand, D.F.; Gan, S.S. Evidence supporting a role of jasmonic acid in *Arabidopsis* leaf senescence. *Plant Physiol.* **2002**, *128*, 876–884. [CrossRef] [PubMed]
57. Schafer, M.; Meza-Canales, I.D.; Navarro-Quezada, A.; Brutting, C.; Vankova, R.; Baldwin, I.T.; Meldau, S. Cytokinin levels and signaling respond to wounding and the perception of herbivore elicitors in *Nicotiana attenuata*. *J. Integr. Plant Biol.* **2015**, *57*, 198–212. [CrossRef] [PubMed]
58. Stankovic, B.; Davies, E. Intercellular communication in plants: Electrical stimulation of proteinase inhibitor gene expression in tomato. *Planta* **1997**, *202*, 402–406. [CrossRef]
59. Malone, M.; Alarcon, J.J.; Palumbo, L. An hydraulic interpretation of rapid, long-distance wound signaling in the tomato. *Planta* **1994**, *193*, 181–185. [CrossRef]

60. Zimmermann, M.R.; Maischak, H.; Mithofer, A.; Boland, W.; Felle, H.H. System potentials, a novel electrical long-distance apoplastic signal in plants, induced by wounding. *Plant Physiol.* **2009**, *149*, 1593–1600. [CrossRef] [PubMed]
61. Mousavi, S.A.R.; Chauvin, A.; Pascaud, F.; Kellenberger, S.; Farmer, E.E. GLUTAMATE RECEPTOR-LIKE genes mediate leaf-to-leaf wound signalling. *Nature* **2013**, *500*, 422–426. [CrossRef] [PubMed]
62. Howe, G.A.; Jander, G. Plant immunity to insect herbivores. *Ann. Rev. Plant Biol.* **2008**, *59*, 41–66. [CrossRef] [PubMed]
63. Schmelz, E.A.; Engelberth, J.; Alborn, H.T.; Tumlinson, J.H.; Teal, P.E.A. Phytohormone-based activity mapping of insect herbivore-produced elicitors. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 653–657. [CrossRef] [PubMed]
64. Xu, S.Q.; Zhou, W.W.; Pottinger, S.; Baldwin, I.T. Herbivore associated elicitor-induced defences are highly specific among closely related Nicotiana species. *BMC Plant Biol.* **2015**. [CrossRef] [PubMed]
65. Koo, A.J.K.; Gao, X.L.; Jones, A.D.; Howe, G.A. A rapid wound signal activates the systemic synthesis of bioactive jasmonates in *Arabidopsis*. *Plant J.* **2009**, *59*, 974–986. [CrossRef] [PubMed]
66. Glauser, G.; Grata, E.; Dubugnon, L.; Rudaz, S.; Farmer, E.E.; Wolfender, J.L. Spatial and temporal dynamics of jasmonate synthesis and accumulation in *Arabidopsis* in response to wounding. *J. Biol. Chem.* **2008**, *283*, 16400–16407. [CrossRef] [PubMed]
67. Wu, J.Q.; Hettenhausen, C.; Meldau, S.; Baldwin, I.T. Herbivory rapidly activates MAPK signaling in attacked and unattacked leaf regions but not between leaves of *Nicotiana attenuata*. *Plant Cell* **2007**, *19*, 1096–1122. [CrossRef] [PubMed]
68. Hause, B.; Stenzel, I.; Miersch, O.; Maucher, H.; Kramell, R.; Ziegler, J.; Wasternack, C. Tissue-specific oxylipin signature of tomato flowers: Allene oxide cyclase is highly expressed in distinct flower organs and vascular bundles. *Plant J.* **2000**, *24*, 113–126. [CrossRef] [PubMed]
69. Orians, C.M.; Pomerleau, J.; Ricco, R. Vascular architecture generates fine scale variation in systemic induction of proteinase inhibitors in tomato. *J. Chem. Ecol.* **2000**, *26*, 471–485. [CrossRef]
70. Farmer, E.E.; Gasperini, D.; Acosta, I.F. The squeeze cell hypothesis for the activation of jasmonate synthesis in response to wounding. *New Phytol.* **2014**, *204*, 282–288. [CrossRef] [PubMed]
71. Gasperini, D.; Chetelat, A.; Acosta, I.F.; Goossens, J.; Pauwels, L.; Goossens, A.; Dreos, R.; Alfonso, E.; Farmer, E.E. Multilayered organization of jasmonate signalling in the regulation of root growth. *PLoS Genet.* **2015**. [CrossRef] [PubMed]
72. Erb, M.; Gordon-Weeks, R.; Flors, V.; Camanes, G.; Turlings, T.C.; Ton, J. Belowground ABA boosts aboveground production of DIMBOA and primes induction of chlorogenic acid in maize. *Plant Signal. Behav.* **2009**, *4*, 636–638. [CrossRef] [PubMed]
73. Van Dam, N.M.; Oomen, M.W. Root and shoot jasmonic acid applications differentially affect leaf chemistry and herbivore growth. *Plant Signal. Behav.* **2008**, *3*, 91–98. [CrossRef] [PubMed]
74. Tytgat, T.O.G.; Verhoeven, K.J.F.; Jansen, J.J.; Raaijmakers, C.E.; Bakx-Schotman, T.; McIntyre, L.M.; van der Putten, W.H.; Biere, A.; van Dam, N.M. Plants know where it hurts: Root and shoot jasmonic acid induction elicit differential responses in *Brassica olracea*. *PloS ONE* **2013**, *8*, e65502. [CrossRef]
75. Baldwin, I.T.; Schmelz, E.A.; Ohnmeiss, T.E. Wound-induced changes in root and shoot jasmonic acid pools correlate with induced nicotine synthesis in Nicotiana-sylvestris spegazzini and comes. *J. Chem. Ecol.* **1994**, *20*, 2139–2157. [CrossRef] [PubMed]
76. Fragoso, V.; Rothe, E.; Baldwin, I.T.; Kim, S.G. Root jasmonic acid synthesis and perception regulate folivore-induced shoot metabolites and increase *Nicotiana attenuata* resistance. *New Phytol.* **2014**, *202*, 1335–1345. [CrossRef] [PubMed]
77. Dyer, M.I.; Acra, M.A.; Wang, G.M.; Coleman, D.C.; Freckman, D.W.; Mcnaughton, S.J.; Strain, B.R. Source-sink carbon relations in 2 Panicum-coloratum ecotypes in response to herbivory. *Ecology* **1991**, *72*, 1472–1483. [CrossRef]
78. Briske, D.D.; Boutton, T.W.; Wang, Z. Contribution of flexible allocation priorities to herbivory tolerance in C-4 perennial grasses: An evaluation with C-13 labeling. *Oecologia* **1996**, *105*, 151–159. [CrossRef]
79. Holland, J.N.; Cheng, W.X.; Crossley, D.A. Herbivore-induced changes in plant carbon allocation: Assessment of below-ground C fluxes using carbon-14. *Oecologia* **1996**, *107*, 87–94. [CrossRef]

80. Babst, B.A.; Ferrieri, R.A.; Gray, D.W.; Lerdau, M.; Schlyer, D.J.; Schueller, M.; Thorpe, M.R.; Orians, C.M. Jasmonic acid induces rapid changes in carbon transport and partitioning in Populus. *New Phytol.* **2005**, *167*, 63–72. [CrossRef] [PubMed]
81. Babst, B.A.; Ferrieri, R.A.; Thorpe, M.R.; Orians, C.M. Lymantria dispar herbivory induces rapid changes in carbon transport and partitioning in *Populus nigra*. *Entomol. Exp. Appl.* **2008**, *128*, 117–125. [CrossRef]
82. Gomez, S.; Steinbrenner, A.D.; Osorio, S.; Schueller, M.; Ferrieri, R.A.; Fernie, A.R.; Orians, C.M. From shoots to roots: Transport and metabolic changes in tomato after simulated feeding by a specialist Lepidopteran. *Entomol. Exp. Appl.* **2012**, *144*, 101–111. [CrossRef]
83. Diezel, C.; Kessler, D.; Baldwin, I.T. Pithy protection: *Nicotiana attenuata*'s jasmonic acid-mediated defenses are required to resist stem-boring Weevil larvae. *Plant Physiol.* **2011**, *155*, 1936–1946. [CrossRef] [PubMed]
84. Kessler, D.; Diezel, C.; Baldwin, I.T. Changing pollinators as a means of escaping herbivores. *Curr. Biol.* **2010**, *20*, 237–242. [CrossRef] [PubMed]
85. Li, L.; Zhao, Y.F.; McCaig, B.C.; Wingerd, B.A.; Wang, J.H.; Whalon, M.E.; Pichersky, E.; Howe, G.A. The tomato homolog of CORONATINE-INSENSITIVE1 is required for the maternal control of seed maturation, jasmonate-signaled defense responses, and glandular trichome development. *Plant Cell* **2004**, *16*, 783–783. [CrossRef] [PubMed]
86. Stitz, M.; Hartl, M.; Baldwin, I.T.; Gaquerel, E. Jasmonoyl-L-isoleucine coordinates metabolic networks required for anthesis and floral attractant emission in wild tobacco (*Nicotiana attenuata*). *Plant Cell* **2014**, *26*, 3964–3983. [CrossRef] [PubMed]
87. Tian, D.L.; Tooker, J.; Peiffer, M.; Chung, S.H.; Felton, G.W. Role of trichomes in defense against herbivores: Comparison of herbivore response to woolly and hairless trichome mutants in tomato (*Solanum lycopersicum*). *Planta* **2012**, *236*, 1053–1066. [CrossRef] [PubMed]
88. Tissier, A. Glandular trichomes: What comes after expressed sequence tags? *Plant J.* **2012**, *70*, 51–68. [CrossRef] [PubMed]
89. Weinhold, A.; Baldwin, I.T. Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7855–7859. [CrossRef] [PubMed]
90. Kang, J.H.; Liu, G.H.; Shi, F.; Jones, A.D.; Beaudry, R.M.; Howe, G.A. The tomato odorless-2 mutant is defective in trichome-based production of diverse specialized metabolites and broad-spectrum resistance to insect herbivores. *Plant Physiol.* **2010**, *154*, 262–272. [CrossRef] [PubMed]
91. Yoshida, Y.; Sano, R.; Wada, T.; Takabayashi, J.; Okada, K. Jasmonic acid control of GLABRA3 links inducible defense and trichome patterning in *Arabidopsis*. *Development* **2009**, *136*, 1039–1048. [CrossRef] [PubMed]
92. Qi, T.C.; Song, S.S.; Ren, Q.C.; Wu, D.W.; Huang, H.; Chen, Y.; Fan, M.; Peng, W.; Ren, C.M.; Xie, D.X. The jasmonate-ZIM-domain proteins interact with the WD-repeat/bHLH/MYB complexes to regulate jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*. *Plant Cell* **2011**, *23*, 1795–1814. [CrossRef] [PubMed]
93. Traw, M.B.; Bergelson, J. Interactive effects of jasmonic acid, salicylic acid, and gibberellin on induction of trichomes in *Arabidopsis*. *Plant Physiol.* **2003**, *133*, 1367–1375. [CrossRef] [PubMed]
94. Katsir, L.; Schilmiller, A.L.; Staswick, P.E.; He, S.Y.; Howe, G.A. COI1 is a critical component of a receptor for jasmonate and the bacterial virulence factor coronatine. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 7100–7105. [CrossRef] [PubMed]
95. Browse, J. The power of mutants for investigating jasmonate biosynthesis and signaling. *Phytochemistry* **2009**, *70*, 1539–1546. [CrossRef] [PubMed]
96. Yao, K.N.; Deluca, V.; Brisson, N. Creation of a metabolic sink for tryptophan alters the phenylpropanoid pathway and the susceptibility of potato to Phytophthora-infestans. *Plant Cell* **1995**, *7*, 1787–1799. [CrossRef] [PubMed]
97. Li, L.; van Eck, J. Metabolic engineering of carotenoid accumulation by creating a metabolic sink. *Transgenic Res.* **2007**, *16*, 581–585. [CrossRef] [PubMed]
98. Qin, G.J.; Gu, H.Y.; Zhao, Y.D.; Ma, Z.Q.; Shi, G.L.; Yang, Y.; Pichersky, E.; Chen, H.D.; Liu, M.H.; Chen, Z. L.; *et al*. An indole-3-acetic acid carboxyl methyltransferase regulates *Arabidopsis* leaf development. *Plant Cell* **2005**, *17*, 2693–2704. [CrossRef] [PubMed]
99. Varbanova, M.; Yamaguchi, S.; Yang, Y.; McKelvey, K.; Hanada, A.; Borochov, R.; Yu, F.; Jikumaru, Y.; Ross, J.; Cortes, D.; *et al*. Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. *Plant Cell* **2007**, *19*, 32–45. [CrossRef] [PubMed]

100. Tieman, D.; Zeigler, M.; Schmelz, E.; Taylor, M.G.; Rushing, S.; Jones, J.B.; Klee, H.J. Functional analysis of a tomato salicylic acid methyl transferase and its role in synthesis of the flavor volatile methyl salicylate. *Plant J.* **2010**, *62*, 113–123. [CrossRef] [PubMed]
101. Stitz, M.; Gase, K.; Baldwin, I.T.; Gaquerel, E. Ectopic expression of AtJMT in *Nicotiana attenuata*: creating a metabolic sink has tissue-specific consequences for the jasmonate metabolic network and silences downstream gene expression. *Plant Physiol.* **2011**, *157*, 341–354. [CrossRef] [PubMed]
102. Stitz, M.; Baldwin, I.T.; Gaquerel, E. Diverting the flux of the JA pathway in *Nicotiana attenuata* compromises the plant's defense metabolism and fitness in nature and glasshouse. *PloS ONE* **2011**, *6*, e25925. [CrossRef] [PubMed]
103. Seo, H.S.; Song, J.T.; Cheong, J.J.; Lee, Y.H.; Lee, Y.W.; Hwang, I.; Lee, J.S.; Choi, Y.D. Jasmonic acid carboxyl methyltransferase: A key enzyme for jasmonate-regulated plant responses. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4788–4793. [CrossRef] [PubMed]
104. Potenza, C.; Aleman, L.; Sengupta-Gopalan, C. Targeting transgene expression in research, agricultural, and environmental applications: Promoters used in plant transformation. *In Vitro Cell. Dev. Biol. Plant* **2004**, *40*, 1–22. [CrossRef]
105. Qin, H.; Gu, Q.; Zhang, J.L.; Sun, L.; Kuppu, S.; Zhang, Y.Z.; Burow, M.; Payton, P.; Blumwald, E.; Zhang, H. Regulated expression of an isopentenyltransferase gene (IPT) in peanut significantly improves drought tolerance and increases yield under field conditions. *Plant Cell Physiol.* **2011**, *52*, 1904–1914. [CrossRef] [PubMed]
106. Xiao, B.Z.; Chen, X.; Xiang, C.B.; Tang, N.; Zhang, Q.F.; Xiong, L.Z. Evaluation of seven function-known candidate genes for their effects on improving drought resistance of transgenic rice under field conditions. *Mol. Plant* **2009**, *2*, 73–83. [CrossRef] [PubMed]
107. Corrado, G.; Karali, M. Inducible gene expression systems and plant biotechnology. *Biotechnol. Adv.* **2009**, *27*, 733–743. [CrossRef] [PubMed]
108. Schafer, M.; Brutting, C.; Gase, K.; Reichelt, M.; Baldwin, I.; Meldau, S. "Real time" genetic manipulation: A new tool for ecological field studies. *Plant J.* **2013**, *76*, 506–518. [CrossRef] [PubMed]
109. Li, L.; Li, C.; Lee, G.I.; Howe, G.A. Distinct roles for jasmonate synthesis and action in the systemic wound response of tomato. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6416–6421. [CrossRef] [PubMed]
110. Fragoso, V.; Goddard, H.; Baldwin, I.T.; Kim, S.G. A simple and efficient micrografting method for stably transformed *Nicotiana attenuata* plants to examine shoot-root signaling. *Plant Methods* **2011**. [CrossRef] [PubMed]
111. Matthes, M.C.; Pickett, J.A.; Napier, J.A. Natural variation in responsiveness of *Arabidopsis thaliana* to methyl jasmonate is developmentally regulated. *Planta* **2008**, *228*, 1021–1028. [CrossRef] [PubMed]
112. Ahmad, S.; van Hulten, M.; Martin, J.; Pieterse, C.M.J.; van Wees, S.C.M.; Ton, J. Genetic dissection of basal defence responsiveness in accessions of *Arabidopsis thaliana*. *Plant Cell Environ.* **2011**, *34*, 1191–1206. [CrossRef] [PubMed]
113. Kliebenstein, D.J.; Figuth, A.; Mitchell-Olds, T. Genetic architecture of plastic methyl jasmonate responses in *Arabidopsis thaliana*. *Genetics* **2002**, *161*, 1685–1696. [PubMed]
114. Gaquerel, E.; Stitz, M.; Kallenbach, M.; Baldwin, I.T. Jasmonate signaling in the field, part II: Insect-guided characterization of genetic variations in jasmonate-dependent defenses of transgenic and natural *Nicotiana attenuata* populations. *Methods Mol. Biol.* **2013**, *1011*, 97–109. [PubMed]
115. Bahulikar, R.A.; Stanculescu, D.; Preston, C.A.; Baldwin, I.T. ISSR and AFLP analysis of the temporal and spatial population structure of the post-fire annual, *Nicotiana attenuata*, in SW Utah. *BMC Ecol.* **2004**. [CrossRef] [PubMed]
116. Baldwin, I.T.; Morse, L. Up in smoke: II. germination of *Nicotiana attenuata* in response to smoke-derived cues and nutrients in burned and unburned soils. *J. Chem. Ecol.* **1994**, *20*, 2373–2391. [CrossRef] [PubMed]
117. Baldwin, I.T.; Staszak-Kozinski, L.; Davidson, R. Up in smoke: I. Smoke-derived germination cues for postfire annual, *Nicotiana attenuata* Torr. Ex. Watson. *J. Chem. Ecol.* **1994**, *20*, 2345–2371. [CrossRef] [PubMed]
118. Kessler, A.; Halitschke, R.; Baldwin, I.T. Silencing the jasmonate cascade: Induced plant defenses and insect populations. *Science* **2004**, *305*, 665–668. [CrossRef] [PubMed]
119. Li, X.C.; Schuler, M.A.; Berenbaum, M.R. Jasmonate and salicylate induce expression of herbivore cytochrome P450 genes. *Nature* **2002**, *419*, 712–715. [CrossRef] [PubMed]

120. Machado, R.A.R.; Ferrieri, A.P.; Robert, C.A.M.; Glauser, G.; Kallenbach, M.; Baldwin, I.T.; Erb, M. Leaf-herbivore attack reduces carbon reserves and regrowth from the roots via jasmonate and auxin signaling. *New Phytol.* **2013**, *200*, 1234–1246. [CrossRef] [PubMed]
121. Steppuhn, A.; Schuman, M.C.; Baldwin, I.T. Silencing jasmonate signalling and jasmonate-mediated defences reveals different survival strategies between two *Nicotiana attenuata* accessions. *Mol. Ecol.* **2008**, *17*, 3717–3732. [CrossRef] [PubMed]
122. Wu, J.Q.; Hettenhausen, C.; Schuman, M.C.; Baldwin, I.T. A comparison of two *Nicotiana attenuata* accessions reveals large differences in signaling induced by oral secretions of the specialist herbivore *Manduca sexta*. *Plant Physiol.* **2008**, *146*, 927–939. [CrossRef] [PubMed]
123. Johal, G.S.; Balint-Kurti, P.; Well, C.F. Mining and harnessing natural variation: A little MAGIC. *Crop Sci.* **2008**, *48*, 2066–2073. [CrossRef]
124. Chan, E.K.F.; Rowe, H.C.; Corwin, J.A.; Joseph, B.; Kliebenstein, D.J. Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLoS Biol.* **2011**. [CrossRef] [PubMed]
125. Riedelsheimer, C.; Lisec, J.; Czedik-Eysenberg, A.; Sulpice, R.; Flis, A.; Grieder, C.; Altmann, T.; Stitt, M.; Willmitzer, L.; Melchinger, A.E. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 8872–8877. [CrossRef] [PubMed]
126. Toubiana, D.; Batushansky, A.; Tzfadia, O.; Scossa, F.; Khan, A.; Barak, S.; Zamir, D.; Fernie, A.R.; Nikoloski, Z.; Fait, A. Combined correlation-based network and mQTL analyses efficiently identified loci for branched-chain amino acid, serine to threonine, and proline metabolism in tomato seeds. *Plant J.* **2015**, *81*, 121–133. [CrossRef] [PubMed]
127. Alseekh, S.; Tohge, T.; Wendenberg, R.; Scossa, F.; Omranian, N.; Li, J.; Kleessen, S.; Giavalisco, P.; Pleban, T.; Mueller-Roeber, B.; *et al.* Identification and mode of inheritance of quantitative trait loci for decondary metabolite sbundance in tomato. *Plant Cell* **2015**, *27*, 485–512. [CrossRef] [PubMed]
128. Matsuda, F.; Okazaki, Y.; Oikawa, A.; Kusano, M.; Nakabayashi, R.; Kikuchi, J.; Yonemaru, J.I.; Ebana, K.; Yano, M.; Saito, K. Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. *Plant J.* **2012**, *70*, 624–636. [CrossRef] [PubMed]
129. Gong, L.; Chen, W.; Gao, Y.Q.; Liu, X.Q.; Zhang, H.Y.; Xu, C.G.; Yu, S.B.; Zhang, Q.F.; Luo, J. Genetic analysis of the metabolome exemplified using a rice population. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20320–20325. [CrossRef] [PubMed]
130. Kusano, M.; Yang, Z.G.; Okazaki, Y.; Nakabayashi, R.; Fukushima, A.; Saito, K. Using metabolomic approaches to explore chemical diversity in rice. *Mol. Plant* **2015**, *8*, 58–67. [CrossRef] [PubMed]
131. Wen, W.W.; Li, D.; Li, X.; Gao, Y.Q.; Li, W.Q.; Li, H.H.; Liu, J.; Liu, H.J.; Chen, W.; Luo, J.; *et al.* Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature Commun.* **2014**. [CrossRef] [PubMed]
132. Agerbirk, N.; Olsen, C.E.; Heimes, C.; Christensen, S.; Bak, S.; Hauser, T.P. Multiple hydroxyphenethyl glucosinolate isomers and their tandem mass spectrometric distinction in a geographically structured polymorphism in the crucifer Barbarea vulgaris. *Phytochemistry* **2015**, *115*, 130–142. [CrossRef] [PubMed]

Manuscript III

Illuminating a plant's tissue-specific metabolic diversity using computational metabolomics and information theory

Dapeng Li, Sven Heiling, Ian T. Baldwin, and Emmanuel Gaquerel

Published in Proceedings of the National Academy of Sciences of the United States of America 2016, 113(47), E7610-E7618.

Illuminating a plant's tissue-specific metabolic diversity using computational metabolomics and information theory

Dapeng Li[a], Sven Heiling[a], Ian T. Baldwin[a] and Emmanuel Gaquerel[a,b,c]

[a]Max Planck Institute for Chemical Ecology, Department of Molecular Ecology, Jena, Germany,
[b]Centre for Organismal Studies, University of Heidelberg, Im Neuenheimer Feld 360, Heidelberg, Germany
[c]Corresponding author: emmanuel.gaquerel@cos.uni-heidelberg.de

E-mail addresses:
DL, dli@ice.mpg.de; SH, sheiling@ice.mpg.de; ITB, baldwin@ice.mpg.de; EG, emmanuel.gaquerel@cos.uni-heidelberg.de

Abstract (241 words)

Secondary metabolite diversity is considered an important fitness determinant for plants' biotic and abiotic interactions in nature. This diversity can be examined in two dimensions. The first one considers metabolite diversity across plant species. A second way of looking at this diversity is by considering the tissue-specific localization of pathways underlying secondary metabolism within-a-plant. While these cross-tissue metabolite variations are increasingly regarded as important readouts of tissue-level gene function and regulatory processes, these have rarely been comprehensively explored by non-targeted metabolomics. As such, important questions have remained superficially addressed. For instance, which tissues exhibit prevalent signatures of metabolic specialization? And reciprocally which metabolites contribute most to this tissue-specialization in contrast to those exhibiting housekeeping characteristics? Here we explore tissue-level metabolic specialization in Nicotiana attenuata, an ecological model with rich secondary metabolism by combining tissue-wide non-targeted mass spectral data acquisition, information theory analysis, and MS/MS molecular networks. This analysis was conducted for two different methanolic extracts of 14 tissues and deconvoluted 895 non-redundant MS/MS spectra. Using information theory analysis, anthers were found to harbor the most specialized metabolome and, through MS/MS molecular networks, most unique metabolites of anthers and other tissues were annotated. Tissue-metabolite association maps were used to predict tissue-specific gene functions. Predictions for the function of two UDP-glycosyltransferases in flavonoid metabolism were confirmed by virus-induced gene-silencing. The present workflow allows biologists to amortize the vast amount of data produced by modern MS instrumentation in their quest to understand gene function.

Keywords: Secondary metabolism, Mass Spectrometry, Data-independent MS/MS, Metabolomics, Molecular Networks, Information theory.

INTRODUCTION

Plants are elegant synthetic chemists making use of their metabolic prowess to produce complex blends of chemicals. Commonly quoted estimates state that plants produce somewhere in the order of 200,000 chemical structures. Secondary metabolites, also referred to as specialized metabolites or natural products, contribute to the largest fraction of this structural diversity. Compared with their counterparts in central metabolism (primary metabolites), secondary metabolite groups have diversified to the extreme in plant lineages, likely as a result of the multiple ecological roles they fulfill[1]. The high degree of plasticity of secondary metabolism pathways is consistent with the existence of large families of metabolism-related genes such as cytochrome P450s and UDP-glycosyltransferases in plant genomes that can create structural and chemical modifications almost without limits. The majority of metabolic gene functions remains however unknown either because the metabolites that they produce are unknown or significant associations remain to be identified between the expression of specific metabolic genes and characterized metabolic groups.

The biosynthesis of particular secondary metabolites or of complete metabolic groups are frequently taxonomically restricted[2]. For this reason, certain secondary metabolite classes have been used as signature characters for biochemical investigation of specific plant families, for instance quinolizidine alkaloids for Fabaceae[3], tropane and steroidal alkaloids for Solanaceae[4] and iridoids for Lamiaceae[5]. Another way of looking at plant secondary metabolism diversity is to consider the precise tissue-specific localization of pathways responsible for their production. Compositional differences are, for instance, readily apparent across floral tissues that produce metabolic blends very different from their vegetative counterparts[6]. In the most extreme cases, the accumulation of secondary metabolites can be restricted to specific cell types. This is frequently the case for plant defense metabolites that are produced in specialized tissues/cell types as a mean of minimizing autoxicity reactions in the surrounding tissues and/or of maximizing the defensive function of these metabolites towards aggressors that attack in a spatially-specific manner[7,8]. A better exploration of tissue-level metabolic specialization is therefore particularly helpful in understanding the contribution of a given tissue to an organism's fitness.

From a mechanistic standpoint, the accumulation of secondary metabolites in a

given tissue requires the spatial-temporal coordination of a vast array of cellular processes in which systems controlling biosynthesis, storage and degradation are of central importance. Regulatory mechanisms coordinating these processes are only beginning to be uncovered for some model metabolic pathways such as those of the family of glucosinolates in Brassicacae[9]. Coexpression analysis using information about gene and secondary metabolite cross-tissue expression patterns has been successfully applied to infer biosynthetic genes in secondary metabolism[4,10,11]. In Arabidopsis, several omics-based tissue atlases (for instance for gene expression, alternative splicing, proteome, …) are publicly accessible to conduct such type of analysis[12].

Tissue-level non-targeted metabolomics of downstream metabolic readouts is more challenging to implement. Notably, the potential of mass spectrometry (MS)-based metabolomics and of the large-scale acquisition of tandem MS (MS/MS) spectra for as many metabolites as possible within a metabolic profile is severely constrained by the absence of straightforward classification and visualization pipelines that enable facile pathway interpretations. Metabolite annotation and identification are the obvious bottlenecks that thwart the metabolomics analysis of secondary metabolism[13,14]. Ideally, we need approaches that combine the strengths of state-of-art statistical methods currently emerging from the genomics field with the recent advances in metabolomics data mining, such as the method of MS/MS molecular network which allow unknown metabolites to be readily classified based solely on their fragmentation patterns.

Here we developed a pipeline combining tissue-wide non-targeted MS data acquisition and information theory to mine patterns of tissue-specific structural diversity. With this pipeline, we analyzed a compendium of 14 dissected tissues of *Nicotiana attenuata*, an ecological model for chemically-mediated adaptive traits in the wild. The analysis resulted in the deconvolution of 895 non-redundant MS/MS spectra, of which 565 exhibited preferential tissue-specificity. Using information theory analysis, we asked whether certain tissues exhibited a higher degree of tissue metabolic specialization and which MS/MS data were linked to these patterns. From all this information, tissue-metabolite association maps were created to provide predictions about the tissue-level analysis of gene functions, some of which were tested by gene silencing techniques.

RESULTS

A compendium of MS profiles obtained from isolated N. attenuata tissues.

Probing metabolite compositional changes across plant tissues is currently being recognized as a new means of gaining insights into the function of secondary metabolites and their biosynthesis[6] as well as into the metabolic specialization of particular tissues. Here we isolated 14 tissues from 28 and 50-day old *N. attenuata* plants growing under controlled growth conditions in the glasshouse (**Figure 1a**). Pools of 100 mg tissues were extracted using independent extractions with 80% or 20 % methanol to increase the coverage of the metabolome with polar to semi-polar compounds not efficiently extracted by 80% methanol. We used an optimized ultra-high performance liquid chromatography (UHPLC)-electrospray ionization (ESI)-quadrupole time of flight-mass spectrometry (qTOF-MS) method to analyze the metabolome profiles of these tissues. Identical chromatographic conditions were used for the analysis of these two extraction types as retention time consistency for identical mass features -- mass features being *m/z* signals detected at a given retention by the peak picking method -- is one of the criteria implemented in our bioinformatics workflow. As expected, the two independent extractions types generated largely distinct chromatographic profiles as illustrated for representative chromatograms of sepal, filament and seed (**Figure 1b**).

The data-set **(File S1)** was processed using the R package XCMS using optimized parameters and analyzed by principal component analysis, which confirmed that extensive variations in the composition of mass features exist among the different tissue profiles (**Figure 1c**) (**Figure S1**). The XCMS x PCA processing procedure is a very common one and is, together with *a priori* knowledge annotation of prominent mass features and of in-source fragmentation patterns, frequently considered as the central mining step in most metabolomics studies. Yet, patterns revealed from this type of data-mining provide little to no information with respect to compound diversity among samples. This type of biological interpretation critically requires an analysis at the level of metabolites described by deconvoluted spectra and not at the level of individual mass features, which is what prior work has used.

Creating a multi-tissue indiscriminant MS/MS library for metabolite structural analysis

In order to collect a holistic repertoire of structural information on the metabolic diversity in our tissue compendium, we implemented a tissue-wide analytical pipeline for indiscriminant (data-independent) MS/MS (idMS/MS) analysis. Compared with data-dependent acquisition methods involving the pre-selection of a restricted list of precursor ions for collision induced dissociation (CID) fragmentation, this approach considers for fragmentation analysis all signals within an *m/z* range set as large as possible[15]. In recent years, the idMS/MS technique, sometimes referred to as shotgun or broad-scale MS/MS, has gained considerable interest as an exploratory method for metabolomics measurements. In a previous study, we showed that idMS/MS can be efficiently implemented to most qTOF instruments by running replicated measurements of the same sample using idMS/MS at different CID voltages to maximize fragment coverage[16]. Furthermore, since the idMS/MS method has the disadvantage of being uninformative about precursor-to-fragment relationships, we optimized a computational pipeline based on cross-sample correlation calculations to perform fragment relationship assignments with high confidence[16].

Here, we improved the previous computational pipeline for exploiting cross-tissue metabolic variations to gain statistical power in precursor-to-fragment assignments. Panels of **Figure 2** depict key steps of the pipeline and the large coverage of metabolic profiles achieved by this method. Briefly, for each CID voltage, precursor and fragments relationships were assigned using Pearson Correlation Coefficient (PCC) analysis across all tissues (see Method section). idMS/MS spectra reconstructed at each CID voltage were merged into a composite idMS/MS spectra and some of redundant sub-idMS/MS were grouped simultaneously via the calculation of spectra similarity. Not all putatively redundant sub-idMS/MSs could be merged into respective compound-specific idMS/MSs using the single spectral similarity threshold value applied to the data-set and hence metabolites prone to particularly intense in-source fragmentation frequently produced several idM/MS spectra by the analysis. This was however not an problem for our study as these different sub-idMS/MSs are expected to co-vary across the tissue data-set and to form tight clusters during the structural clustering analysis applied later on in the workflow (**Figure 4**). **Figure 2d** depicts the graphical representation of the idMS/MS

spectrum of a phenolamide metabolite (*N',N"*-caffeoyl,feruloylspermidine)[17] after merging CID voltage-specific idMS/MSs. In this centered network representation, the top node corresponds to the precursor ion ($[M+H]^+$, *m/z* 484.243 @ 506 s) and other nodes to assigned fragments with edge thickness inversely proportional to the *P*-value calculated for a given assignment. Interestingly, due to the in-source fragmentation, sub-idMS/MS of the caffeoylspermidine (*m/z* 308.196) and feruloylspermidine (*m/z* 322.212) fragment could also be constructed at the same retention time resulting in redundant deconvolutions. The deconvolution efficiency was tested by comparing idMS/MS with previously reported MS/MS spectra at an optimized CID voltage for major *N. attenuata* secondary metabolites as well as unknowns (**Figure S2**). The idMS/MS coverage was additionally visualized for an 80% methanol extracted pedicel sample in which 370 idMS/MS spectra were reconstructed (**Figure 2b**). In this visualization, the coloring within retention time heatmap aligned to the profile mode chromatogram reveals that a high density of idMS/MSs were assembled at the regions where intense chromatographic peaks appeared. A idMS/MS coverage view for all tissues under different extraction conditions was additionally constructed using a density plot for 28 tissue profiles (**Figure 2c**). Altogether, the computational pipeline (merging of CID-voltage specific data and partial redundancy filtering) retrieved a library of 895 non-redundant idMS/MSs (**Files S2 and S3**); these were used as the data for all subsequent analyses presented here.

Tissues differ in their degree of metabolic specialization

For a first perspective into tissue metabolic-relationships, we normalized idMS/MS spectra intensities (precursor intensities) using a modified z-score method, termed ZMAD (see Methods) and used hierarchical clustering analysis (HCA) (**Figure S3**). When merging the data-sets obtained from the two extraction procedures, three main clusters appeared from the HCA based on Euclidean distance calculations (**Figure 3a**): one cluster with most non-reproductive tissues of flowers (corolla limb, corolla tube, sepal, pedicel, bud and filament), one with the reproductive parts (anther, nectary, ovary and style and stigma) and a last one with vegetative tissues (leaf, root, stem and seed). Interestingly, tissues that connect reproductive and non-reproductive parts in flowers, namely filaments and stamens, exhibited strongly divergent idMS/MS compositional profiles, demonstrating that relatively fine-scale spatial modulations of

metabolism can be analyzed by this approach. It should be noted that the upstream computational procedure used to deconvolute idMS/MS spectra was performed tissue-wide (and not at the individual tissue-level) and relied on matrix alignment and noise filtering steps to produce an idMS/MS tissue-wide matrix that was of a consistent size (**File S1**). A drawback from this computational approach is that no information about the number of idMS/MSs per tissue is readily available to explore tissue-level metabolic specialization. Intuitively, the presence of few high intensity idMS/MS in a given tissue compared to the average calculated across all tissues could be indicative of high degree of metabolic specialization, while the presence of a large number of average intensity idMS/MSs could reflect a low metabolic specialization. Such interpretations are linked to the frequency distribution of each idMS/MS within the data-set. Information theory, which was pioneered by Shannon in a seminal article in 1948[18], provides the statistical framework to cope with this type of analysis. In defining tissue metabolic diversity and specialization, we therefore considered idMS/MS as symbols, in the sense of information theory, and estimated for each tissue's metabolome its diversity based on the Shannon entropy of its frequency distribution. In other words, tissue-level metabolome specialization was measured as the average specificity of each of its idMS/MS component. Using previously implemented formulae[19], we retrieved values for the following indexes: Hj reflecting the tissue-level idMS/MS diversity and δj for the tissue-level idMS/MS specialization as inferred from the average idMS/MS specificity in the data-set.

Visualizing tissue metabolic profiles in a two-dimensional space using these two indexes as coordinates, revealed a number of interesting patterns (**Figure 3b**). A most obvious one was that tissues significantly vary in their degree of specialization (δj) and diversity (Hj). When extraction types were merged to achieve a more comprehensive view, anthers emerged as the tissue with the least diverse, most specialized metabolome (Hj = 5.16, δj = 1.95). In other words, several idMS/MSs exhibited relative high intensity levels concomitant with low frequency distributions across tissues. Root (Hj = 6.66, δj = 1.49), stem (Hj = 6.91, δj = 1.32) and sepal (Hj = 7.39, δj = 1.38) samples followed anthers in terms of low idMS/MS diversity and mid-to-high idMS/MS specialization. In the case of roots, the relatively high specialization index value retrieved for this tissue was especially supported by idMS/MS spectra collected from the 20% methanol extraction (**Figure S4**). The

signature for low diversity and low specialization detected in seeds (Hj = 6.34, δj = 1.09) was in line with the low density of chromatographic peaks seen for this tissue. Style and stigma (Hj = 6.47, δj = 1.11), filaments (Hj = 7.32, δj = 1.08), ovary (Hj = 7.31, δj = 1.10), corolla tube (Hj = 7.81, δj = 1.12) and pedicel (Hj = 8.17, δj = 1.16) were the tissues exhibiting lowest specialization indexes. The pedicel had the most diverse idMS/MS profile of all tissues analyzed.

Tissue-level differentiations in 17-hydroxygeranyllinalool diterpene glycosides and phenolamines

As a first step towards mining metabolite compositional variations across tissues, we first amortized previous chemical knowledge acquired from *N. attenuata* leaves and evaluated whether the distribution across tissues was differentially modulated at different levels of know secondary metabolic pathways. For this, we selected as case study the 17-hydroxygeranyllinalool diterpene glycosides (HGL-DTGs) pathway that produces abundant acyclic diterpenes with antiherbivore functions[20]. For this metabolic group, we retrieved the corresponding idMS/MS for the most abundant metabolites and investigated cross-tissue modulations as visualized by plotting individual tissue ZMAD normalized values (**Figure S5**). HGL-DTGs can be sub-categorized based on their sugar/malonyl decorations as follows: the precursor molecule (lyciumoside I); core structures with higher-degree of glycosylation but no malonyl groups (nicotianoside III, lyciumoside IV and attenoside); and mono- (nicotianosides IV, Ia and VI) and dimalonylated structures (nicotianosides V, II and VII). Lyciumoside I and lyciumoside IV, its direct rhamnosylation product, were detected in young and photosynthetically active rosette leaves and at lower normalized levels in certain floral organs. The analysis revealed that HGL-DTGs varied significantly among tissues, and the variance was organized by biosynthetic sequence in the pathway. The general trend was that greater tissue-specific variation was found in the downstream steps of the pathway. This was particularly apparent for mono-malonylated HGL-DTGs and suggests an increased translocation from source to sink tissues that increased with HGL-DTG glycosylation and malonylation. Di-malonylated HGL-DTGs were more abundant in certain reproductive organs relative to all other tissues. Another pathway monitoring example is provided for the phenolamide pathway for which an apparent greater specificity towards certain

reproductive organs was detected for polyacylated spermidine conjugates (**Figure S6**).

Large-scale inference of idMS/MS tissue-specificity reveals basic principles of tissue interdependencies

The above descriptions confirmed that tissue-based differentiations are detectable for characterized metabolic pathways. This is consistent with the idea that specific groups of metabolites specifically accumulate in one or several tissues albeit being detectable at lower levels in almost all others tissues. The specificity index of information theory (Sj) of a given idMS/MS serving for δj calculation tends to be stringent and excludes features exhibiting significant degree of specificity (association) with more than one tissue (**Figure 4a**, middle panel). In an attempt to statistically assess the degree of association of idMS/MS towards one or several tissues, we analyzed idMS/MS expression distribution across tissues using reduction of Kurtosis as developed in Li et *al.*[21]. The Kurtosis analysis measures expression distribution patterns rather than frequencies and skirts the restriction of the number of tissues with which a given idMS/MS can be associated. As such, the method has been found to be highly successful in detecting tissue-specificity from large-scale data. Briefly, idMS/MS spectra that exhibit high tissue-specificity are characterized by a high kurtosis value with either a right- or left-tailed leptokurtic distributions, whereas low tissue-specificity idMS/MS exhibit a low kurtosis with normal distributions (**Figures 4a and 4b**). A total of 595 out of 895 idMS/MS exhibited preferential tissue-associations ($Q<0.05$), the rest of the idMS/MS being considered as non-tissue-associated features. For ease of interpretation, **Figure S7** and **S8** report the statistical support via mapping of Kurtosis Q values and inferred tissue-associations for previously discussed tissue-level differentiations in the HGL-DTG and phenolamide pathways.

To retrieve tissue-idMS/MS specific associations, we defined a tissue relative expression threshold Z ($Z = 2$) through the calculation of a reduction of kurtosis according to the rationale proposed by Li et *al.*[21] (**Figure 4c**) (**Figure S9**). Seeds harbored again the smallest associated metabolome with 99 specifically expressed idMS/MS, followed by stem (158 idMS/MS) and root (278 idMS/MS), while a general trend was that floral organs had the largest numbers of associated idMS/MS. An interesting level of analysis was therefore to look at the % of tissue-specific idMS/MS

compared with non-specific ones per tissue. For instance, a number albeit small (278 idMS/MS), idMS/MS specific to roots, these represented 72.3% of the total detectable root metabolome, indicating the relatively high metabolic specialization of this tissue.

In agreement with the importance of not restricting the analysis only to single tissue-specific idMS/MS and considering different degrees of specifity based on the number of tissues in which a given idMS/MS accumulates, we detected that idMS/MS specifically associated with more than one tissue were highly prevalent (97%) among the tissue-specific idMS/MS pool (**Figure S10**). The relative strength of the metabolic interdependencies between two tissues based on the number of shared tissue-specific idMS/MS was scored using the Jaccard index (**Figure 4d**). Clustering based on this score again supported the fact that vegetative tissues such as leaf, stem and root cluster apart from floral counterparts in terms of secondary metabolite profiles. The "floral" cluster subdivided into three smaller clusters: one with tight connections between tissues not directly involved in reproductive tissues (besides the complete bud), one comprising tissues with mostly reproductive functions (filament, style, ovary but also the nectary) and finally anthers. The individualized positioning of anthers in this clustering analysis is in agreement with the information theory specialization signature detected in this tissue, as discussed above. Furthermore, a tissue network based on shared tissue-specific idMS/MS was also constructed based on Ochiai similarities between tissue-pairs to detect the relative contribution of the two extraction methods for the above clustering outcome (**Figure S11**).

MS/MS structural analysis of idMS/MS associations

Examples of annotated tissue-specific idMS/MS spectra shared by different tissues are presented in **File S3**. Metabolite annotation remains a bottleneck in metabolomics studies as public spectral databases are poorly populated with plant specialized metabolites, many of them being taxa- and frequently species-specific. The MS/MS molecular network method pioneered by the Dorrestein lab (citation needed) circumvents the limitation of spectral databases via the analysis of within-a-dataset MS/MS similarities to accelerate hypothesis generation about the identity of unknown MS/MS[22]. This approach also has the advantage of being amenable to the visualization of putative biochemical relationships among metabolites corresponding to highly similar MS/MS spectra[16]. In a recent study, we improved the

scoring and classification of MS/MS similarities for plant secondary metabolites notably by the implementation of a bi-clustering method that detects possible compound familial groupings according to fragment and neutral loss (NL)-based similarities[16]. Applying this method to the total pool of idMS/MSs from the present study resulted in the formation of 9 modules within which idMS/MSs are expected to share high structural similarities (**Figure 5a**). Modules 4, 6, 7, 8 and 9 largely corresponded to previously identified compound classes: flavonoid glycosides, phenolics, HGL-DTGs, acyl sugars and nicotine and polyamine derivatives, respectively. As expected, uncharacterized metabolites likely belonging to these groups and not previously thoroughly investigated were also detected as sub-idMS/MSs that did not merge during the redundancy filtering step. A comprehensive view is provided in **File S3** that concatenates MS/MS spectral content and NLs, their clustering and association with tissues. A critical consideration was whether tissues differ in their module relative composition as depicted in the stacked bar chart of **Figure 5b**. For instance, it is clearly visible that complete *O*-acylsugar metabolism (M8) is absent from anthers and the style, that the stem and seeds lack HGL-DTG metabolism (M7) or that the flavonoid module (M4) is overrepresented in certain flower tissues. Molecular networks can be constructed for each module to better visualize structural relationships among idMS/MSs. The case of module M4 is presented (**Figure 6**). A sub-part of this flavonoid-enriched module contains *O*-acylsugar type II due to shared NLs with flavonoid glycosides, those two groups are still discriminated according the edge density and by considering the idMS/MS tissue co-expression scores. By simply mapping the relative expression of idMS/MS onto nodes, it is possible to rapidly pinpoint metabolites that are characteristic of a given tissue, for instance the dramatic overrepresentation of kaempferol-o-glucoside (KG) at the limb level (808.684 Z-MAD scaled intensity). Also, the idMS/MS for *m/z* 295.102 specific to anthers and not co-expressed across tissues with any other flavonoid glycosides from module M4 is putatively annotated by our method as a glucose ester with C4 side chains, here depicted as 6-tuliposide B[23]. Additional annotations of the top anther-specific idMS/MS are presented in **File S3**.

Exploring metabolite and gene co-associations across tissues facilitates metabolic gene pathway assignment

In this last section, we illustrate the power of first determining tissue-metabolite associations in generating predictions about the assignment of unknown genes to particular pathways. In the case of a uni-modal regulation (with cross-tissue transport being minimal), the logic behind these predictions is that a gene responsible for the production of given set of metabolites will share maximal tissue-associations with these metabolites. As for gene expression data, we used a RNAseq transcriptome data-set in which tissues and developmental stages largely overlap with those used for metabolomics but which also included treatment responses to account for the fact that certain genes are expressed constitutively at low levels but the metabolites can accumulate without turnover to high levels. Similarly to idMS/MS data, the Kurtosis filtering allowed us to filter out genes with quasi-constant expression and focus on those exhibiting leptokurtic distributions (**Figures S13 and S14**). 37% of the total gene expressed exhibited leptokurtic distributions, i.e. expressed specifically in one or several tissues (**Figure 7a**). Subsequent analysis steps followed those presented above for the analysis of metabolites. Overrepresented gene ontologies (GOs) within the complete set of genes with preferential tissue associations corresponded to general processes such as chloroplast thylakoid activity, monocarboxylic acid biosynthetic process, anion transport and metal ion transport (**Figure S15**). This GO overrepresentation analysis was also conducted on a module-basis for a gene set specifically associated with a given module (**Figures S16 and S17**) and confirmed considerable variations in gene-encoded processes characteristic of certain metabolic groups, such as pollen tube growth (**Figure S17**). Through this approach, it is now possible to target specific metabolic gene families, here UDP-glycosyltransferases, and predict their importance for the metabolic group enriched within a given idMS/MS module (**Figure 7**). For mining this latter gene family, an additional filtering criterion is the presence within the co-associated idMS/MS of NLs corresponding to glucose or rhamnose moieties. Modules 4, 7 and 8 are made of idMS/MS corresponding to glycosylated secondary metabolites and hence enriched in the presence of these latter NLs (**Figure 7b**). We extracted ten members of this gene family which had co-tissue specificities with members of M4 and tested the function of two of them (**Figure S18**), *UDP-glycosyltransferase-A* (*UGT-A*) (PCC = 0.71 with rutin) and *UDP-glycosyltransferase-B* (*UGT-B*) (PCC = 0.71 with quercetin-3-O-glucose) by transient gene silencing using virus-induced gene silencing (VIGS) (**Figure S19**). The

two targeted UDP-glycosyltransferases and their co-association with M4 are shown in **Figure 8a**. Both UDP-glycosyltransferases highly co-expressed with flavonoid glycosides but not with type II *O*-acyl sugars that coexist in the same module. Briefly, when silencing *UGT-A*, a majority of the flavonoid glycosides in flower buds were significant decreased in their accumulations namely KG (kaempferol-3-O-glucose, see **Figure S20** for additional structure elucidation), QG (quercetin-3-O-glucose), KGR (kaempferol-3-O-glucose-rhamnose) and QGG (quercetin-3-O-glucose). On the other hand, silencing *UGT-B* translated into significant decreases in the levels of rhamnose-containing KGR and QGR whereas QGG, QG and KG accumulated to higher levels compared with the empty vector control. This is consistent with the conclusion that *UGT-B* likely controls the rhamnosylation of these flavonoid glycosides and that the higher accumulations of non-rhamnose flavonoid glycosides reflects the metabolic tension existing with the *UGT-A*-mediated glucosylation process. A proposed metabolic scheme of the flavonoid glycosides pathway is presented in **Figure 8c**.

DISCUSSION

In this study, we investigated tissue-level variations in secondary metabolism in an ecological model plant using computational metabolomics and information theory statistics. Information theory has been employed for multivariate data generated in a broad-scope of biological contexts ranging from plant ecology[24] to microbiome diversity[25], but to our knowledge, it has never been used to summarize trends in MS-based metabolomics data. Previous studies identified preferential tissue-based redirectionalities in secondary metabolism, for instance, during the maturation of tomato fruits for which the green, turning and red developmental stages are characterized by rearrangements in pathways related to flavonoids, phenolics and glycoalkaloids[26]. However, to our knowledge, no unbiased metabolomics study, other than that of the AtMetExpress database[6], has been applied with rigorous statistical analysis to such a broad range of tissues as in the present study.

This statistical portfolio revealed that tissues exhibit distinct states of secondary metabolism activity but also that they differ in their degree of specialization. An extracted feature illustrative of the explorative power of this approach was that connecting tissues of flowers such as anthers and filaments, on one hand and that the

corolla limb and tube on the other hand differed dramatically in their metabolite specialization signatures. For the corolla, this contrast highlights the fact that limbs are functionally specialized for attracting and guiding pollinators, and likely require a highly specialized metabolome to fulfill this function. The latter is especially expected in a species such as *N. attenuata* whose main pollinator, the hawkmoth *Manduca sexta*, is also a voracious folivore during its larval stage[27], requiring that the plant critically fine-tunes its blend of secondary metabolites to solve the dilemma imposed by these two contrasted interactions[28,29]. As expected, the green tissues of our data-set display the most prototypic and undifferentiated metabolic profiles as highlighted by both targeted and non-targeted analyses.

N. attenuata is a pioneer plant in post-fire habitats[30,31] and as such represents one of the primary food sources for herbivorous insects[32]. It is well established that the photosynthetically active tissues of this plant mount a very strong specialized metabolic response locally and systemically during biotic challenges such as insect herbivory[33]. It would therefore be very interesting to reassess how the specialization indices readjust during stress adaptation, taking advantage of pre-existing knowledge on anti-herbivory function of many secondary metabolite classes[6,20,34]. Also, the pools of many of these defensive secondary metabolites are rearranged during ontogeny in the form of quantitative gradients established across tissues[20,35]. The optimal defense theory provides a conceptual framework that links these quantitative patterns with the fitness of different tissues for the plant's fitness[36,37]. Even though the developmental stages of multiple tissues would need to be separately analyzed with our analytical approach to evaluate this theory thoroughly several defense-related metabolites exhibited higher relative levels in reproductive tissues than in vegetative counterparts, a central prediction of the optimal defense theory . A last remark concerns the extremely low metabolic diversity detected in seeds, a result which could possibly be due to the fact that most apolar metabolites present in the seed endosperm were poorly recovered with our extraction systems. This speaks for the need to use a more sophisticated combination of extraction and chromatographic systems in future experiments in order to capture the behavior of a broader range of compound classes.

It is tempting to consider that signatures of high metabolic specializations observed for certain tissues correlate with their highly specialized physiological functions. Previous tissue-level -omics analyses in plants and animals are consistent

with the expectation that physiological differentiation is accompanied with qualitative variations of metabolic capacities[38,39]. As previously noted, this claim is difficult to support only with metabolomics due to the sparse knowledge about secondary metabolite biosynthetic schemes. The GO enrichment analysis conducted in this study supports the fact that the Kurtosis-based method is able to discriminate gene signatures involved in the tissue-specialized physiological processes from housekeeping ones -- for instance pollen tube growth **(Figure S17**) --, so it is reasonable to propose that metabolites extracted by the Kurtosis method also reflect tissue-level functions, even if transport processes may obfuscate some of these trends. In this regard, the case of anthers exhibiting a prevalent signature of high idMS/MS specialization, greater than all other reproductive organs, is particularly germane. In line with our observation of the manufacture of specific set of metabolites in this tissue (**File S3**), previous studies have shown that metabolites, notably certain phenolic derivatives, are abundant and highly specific for the tapetum (specialized layer of nutritive cells and source of precursors for the pollen coat within anthers) of anthers and pollen grains[40-42]. The biosynthesis of these phenolic derivatives, with potential roles in pollen coat composition and establishment of fertilization barriers, has been linked to rapid metabolic gene evolution through retroposition and neofunctionalization[43]. In a cross-species study on transcriptome evolutionary divergence, the fastest rates of gene expression divergence and signatures of transcriptome specialization were detected in anthers while the lowest rates of evolution were detected in roots[44]. Our metabolomics study therefore suggests that transcriptome and metabolome specialization may be coupled patterns in anthers, likely as a result of strong reproduction-related selection pressures exerted at this tissue-level. More broadly, it would therefore be very interesting to analyze whether such kinds of metabolic specialization patterns are consistent across species for homologous tissues.

Navigating large data-sets in such a way that knowledge can be more efficiently accessible for hypothesis formulation is one of the challenges that thwarts the routine application of certain omics technologies to non-model systems organisms. In this study, we employed data-independent MS/MS acquisition. This approach, albeit suffering from redundant data collection for certain metabolites prone to intense in-source fragmentation, maximizes the comprehensiveness of fragment data

collection which forms the foundation of such unbiased analysis. The present study also speaks to the power of the previously described molecular network method for plant samples and identifies new directions for its integration with genomics data. The latter is illustrated by our functional studies on UDP-glycosyltransferases and the assignment of two previously uncharacterized genes of *N. attenuata* to the glucosylation and rhamnosylation steps in floral flavonoid glycoside metabolism (**Figure S20**)[45,46]. Importantly, the data-platform generated here can also be mined for additional metabolic gene families (P450, BAHD acyltransferases, etc...).

Population geneticists have educated molecular biologists in how to harness the statistical power of variance arising from natural variation for gene pathway and function elucidations. Much more recently, tissue-level co-expression analysis among genes and metabolites has been shown to be an efficient way forward for gene function analysis in secondary metabolism[6]. This study reinforces the power of applying such approaches in combination with large-scale metabolomics via the use of information theory analysis.

In summary, a major strength of this unique study is that it synergistically combines, using a three-pronged approach, the strengths of *(i)* information theory to capture signatures of diversity and specialization in the data-set, of *(ii)* computational MS to accelerate the structural annotation of the diversity of compounds collected and of *(iii)* experimental gene-silencing to falsify hypotheses regarding metabolic gene functions from metabolomics-transcriptomics integration. A recent breakthrough study on mammals' metabolomes has highlighted the power of metabolomics to predict markers associated with organ specialization in a phylogenetic context[47]. Future directions will make use of genomics resources existing for related species of *N. attenuata* to extend the approach to the diagnosis of gene divergence effects contributing most to tissue-metabolic specialization.

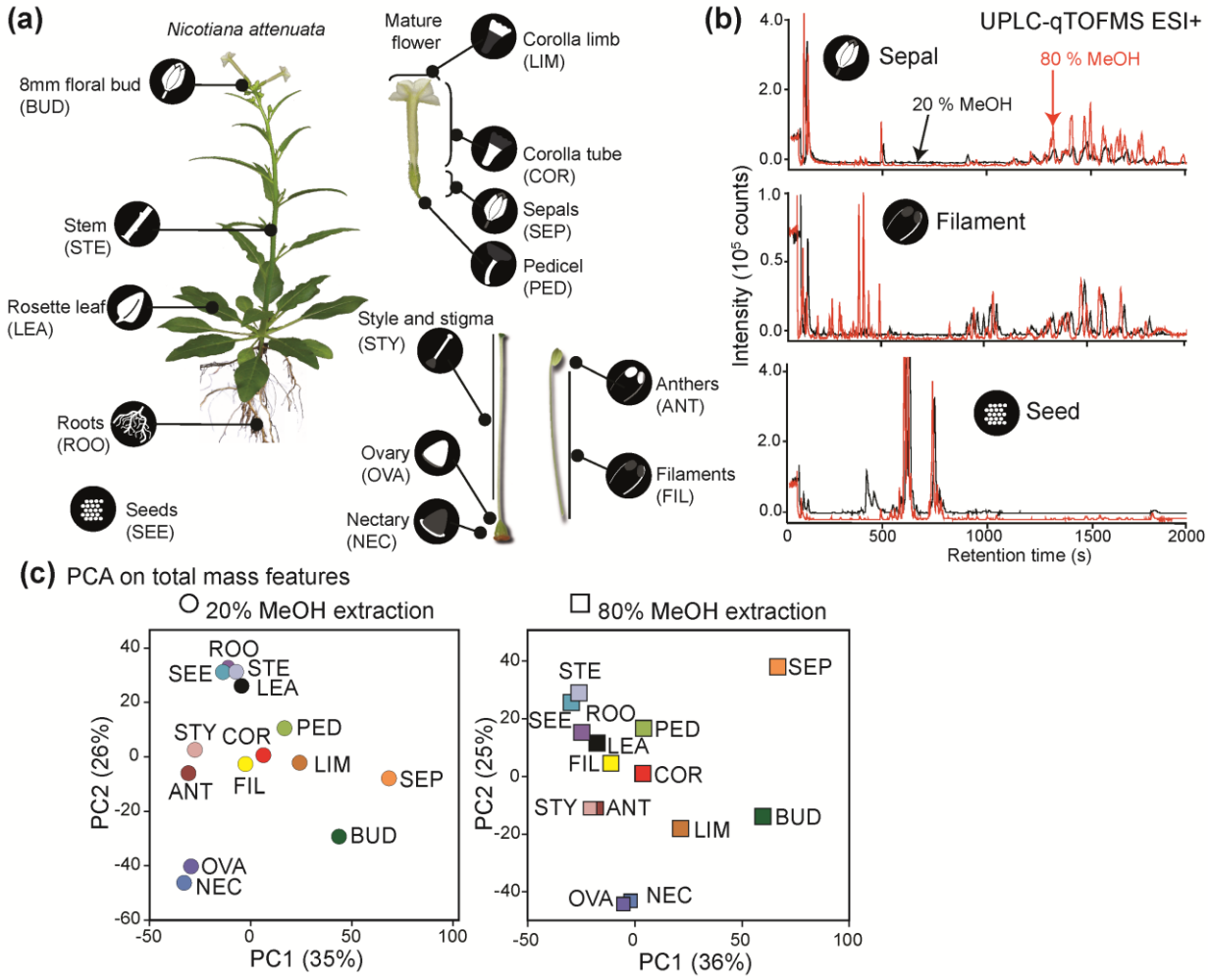

Figure 1. MS-based metabolomics profiling of 14 distinct *N. attenuata* tissues. (a) Tissues were collected and analyzed separately for metabolomic profiling. Detailed explanations of the tissue collection procedure are provided in the Method section. **(b)** UPLC-qTOF profiling by electrospray in positive ionization mode was conducted using previously-optimized chromatographic conditions. Raw data files were converted to non-commercial netCDF format and processed for peak picking and multi-parallel retention time alignment using the R package, XCMS, with optimized parameters. **(c)** Principal component analysis using the autoscaled complete mass feature matrices obtained from processing samples extracted with 20% and 80% methanol. A PCA of the combined data-set is presented in **Figure S1**.

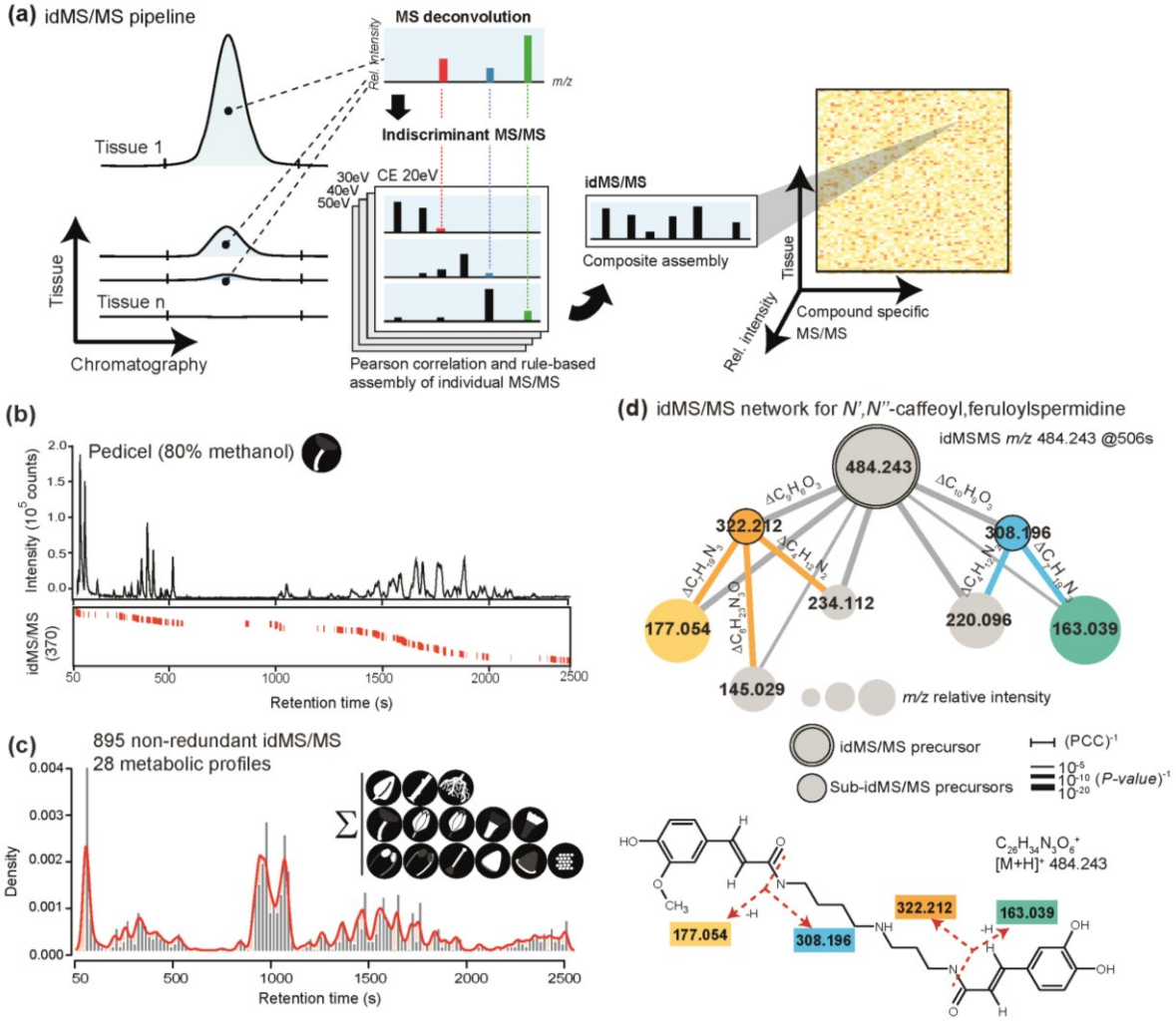

Figure 2. Tissue-wide indiscriminant acquisition and assembly of metabolite MS/MS spectral information. (a) Optimized pipeline to achieve indiscriminant acquisition of MS/MS data from tissue-wide metabolic variations. Indiscriminant MS/MS targets every mass signal within a range of 50 to 1400 *m/z* for fragmentation using 4 increasing collision energies (CE). Each tissue sample is processed using idMS/MS via individual analyses performed at different CID voltage in order to maximize fragment diversity. Information regarding a fragment's assignment to a given precursor mass is lost during indiscriminant MS/MS (idMS/MS) but can be computationally-retrieved for each CID voltage run, based on mathematical and chromatographic correlation analyses (see Method section). The pipeline harnesses the important chemical diversity among tissue samples. Reciprocally, cross-tissue quantitative variation provides the statistical power required for computing high confidence Pearson correlation coefficients (PCC) among the variation in intensity of

precursors and candidate fragments. idMS/MS assembled at each CID voltage for a given precursor *m/z* occurring at a given retention time are then merged together into a composite MS/MS which displays the complete fragment diversity. The resulting output is a three-dimensional entry matrix with non-redundant idMS/MS spectra across tissues and their relative intensities. **(b)** idMS/MS coverage for a representative 80% methanol pedicel sample. The lower heatmap displays the retention time position, aligned to the total ion current chromatogram, of the 370 idMS/MS spectra recorded for this sample. **(c)** Density plot summarizing the idMS/MS coverage across all 14 analyzed tissues denoted by their symbols. Bars represent the relative density of collected idMS/MS for a 9 s retention time window. Red lines represent smoothed density curves. **(d)** Example of the idMS/MS obtained for *m/z* 484.243 @ 506 s corresponds to the $[M+H]^+$ of an *N',N''*-caffeoyl,feruloylspermidine isomer – ' and '' denote that the exact position of the caffeoyl and feruloyl moieties on the spermidine backbone cannot be assigned by MS analysis alone. idMS/MS were also assembled for *m/z* 322.212 and *m/z* 308.196 as these *m/z* signals are already present as in-source ionization derived fragments in the profile mode analysis.

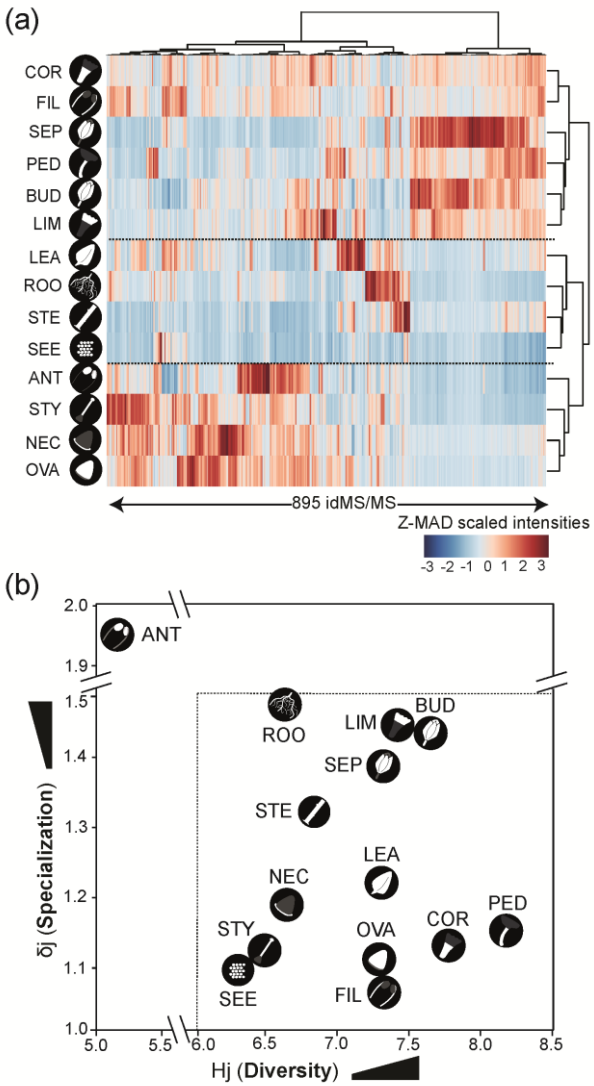

Figure 3. Relative expression of idMS/MS analyzed by information theory reflects tissue-specific metabolome specialization. (a) Hierarchical clustering, using the Euclidean distance as clustering metric, of tissue-specific idMS/MS relative expression profiles. The heatmap coloring depicts the scaled intensities. Z-score-normalized median absolute distances captured the cross-tissue variations for idMS/MS intensity (895 idMS/MS) obtained for each tissue **(b)** Information theory analysis of tissue-level idMS/MS composition specialization (δj) and diversity (Hj) based on idMS/MS cross-tissue distributions is displayed in a two dimensional space to reveal gradients of metabolic specialization. ANT, anthers; BUD, floral bud; COR, corolla tube; FIL, filaments; LEA, rosette leaves; LIM, corolla limb; PED, floral pedicel; ROO, root; SEE, seeds; SEP, floral sepals; STE, stem; STY, floral style.

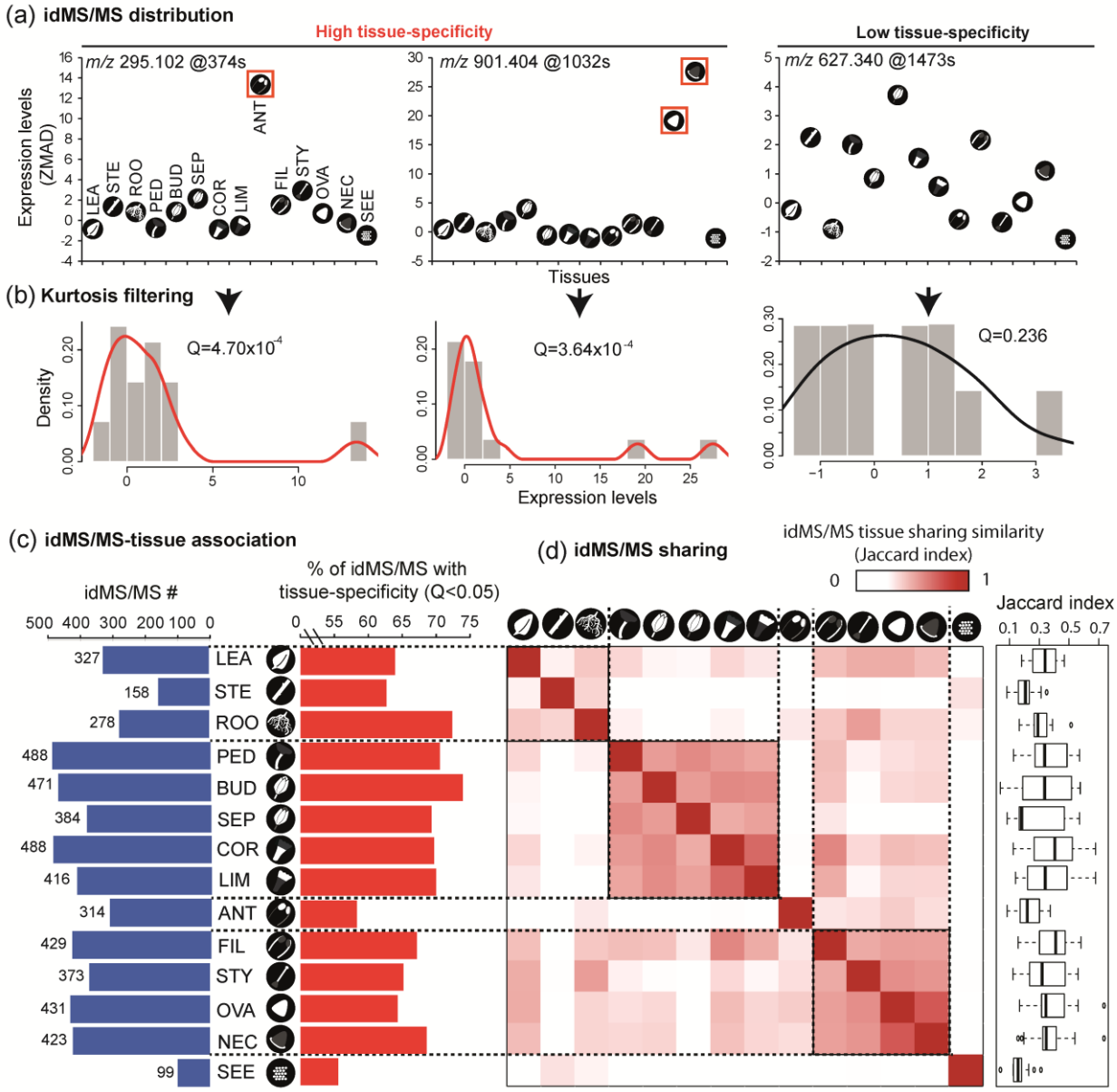

Figure 4. Large-scale analysis of idMS/MS tissue-specificity. (a) Cross-tissue distribution patterns for three idMS/MS examples. Z-score-normalized median absolute distances captured cross-tissue variations for idMS/MS intensity. idMS/MS deconvoluted for *m/z* 295.102 @ 374s and 901.404 @ 1032s revealed clear tissue specificity for one and two tissue types respectively. idMS/MS for *m/z* 627.340 @ 1473s was not associated with a particular tissue. **(b)** The density of intensity levels of each idMSMS across all analyzed tissues is computed and filtered using a reduction of Kurtosis method to determine idMS/MS with significant tissue specificity. **(c)** Left bar chart, number of idMS/MS per tissue using an intensity threshold of 2; right bar chart, % of idMS/MS showing tissue-specificity per tissue. **(d)** Heatmap matrix

visualizing idMSMS sharing among tissues as measured using the Jacquard index. See **File S3** for idMS/MS classifications to main compound classes in *N. attenuata* as obtained by idMS/MS alignments to public libraries and manual curation.

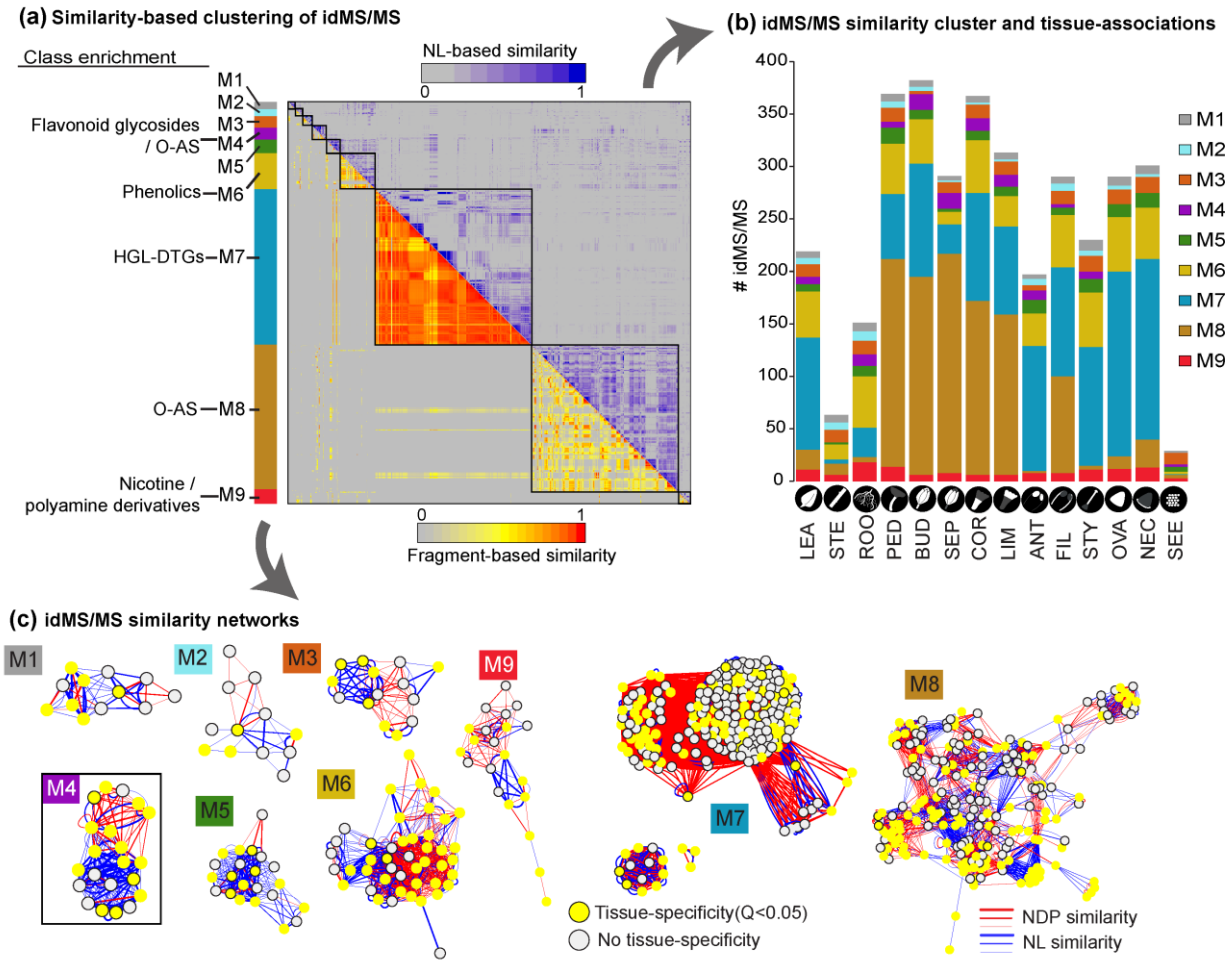

Figure 5. Combination of structural classifications of idMS/MS and tissue-specificity of expression. (a) Biclustering analysis to classify idMS/MS according to structural similarities. The analysis used two scoring methods: one based on shared fragments among spectra, while the other scored shared common neutral losses among spectra. Using biclustering, which favors clustering based on iterative alignments of spectra based on the two scoring methods, produces large modules (M) with structurally-related idMS/MS. Some of these modules were congruent with known compound families while others were comprised of yet unknown or poorly characterized metabolites. Module annotation is reported in **File S3** and idMS/MS intensity distribution in **Figure S12**. **(b)** Relative contribution of each module to the idMS/MS associated with a given tissue. The visualization highlights the complete absence of specific metabolic groups, here corresponding to particular modules, such *O*-acyl sugars (*O*-AS), in anthers. **(c)** Molecular networks constructed for each modules. Nodes represent idMS/MS and edges similarity values based on the two-score types. Tissue-specificity can easily be mapped to the molecular networks.

17-HGL-DTG, 17-hydroxygeranyllinalool glycosides.

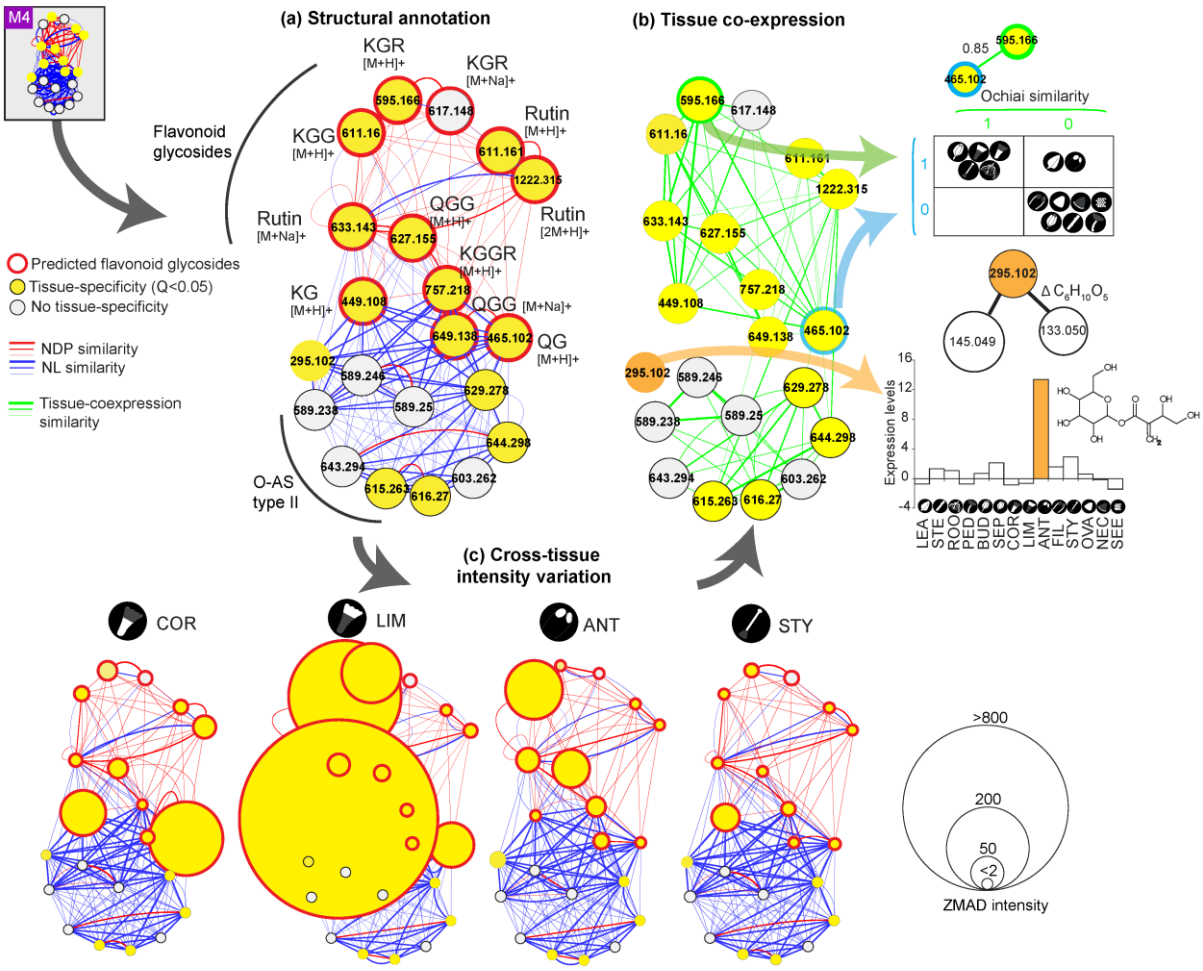

Figure 6. Distribution of a flavonoid-enriched module among different flower parts. (a) Network representation and annotation of module M4 from the bi-clustering analysis. Nodes correspond to idMS/MS spectra and edges to their pairwise similarity as measured according to the fragment (NDP) (> 0.6) and neutral loss (NL) similarity (> 0.6). Many of the spectra correspond to flavonoid glycosides, albeit *O*-acyl sugars of type II are also present due to shared NLs. **(b)** Cross-tissue co-expression (based on the Ochiai score > 0.6) between idMS/MS spectra discriminates flavonoid glycosides from *O*-acyl sugar. The analysis reveals metabolites within the M4 module with high tissue-specificity such as idMS/MS at *m/z* 295.102, predicted to be a tuliposide derivative, which is abundant in anthers (**Figure 4**, panel a). **(C)** Examples of visualization of cross-tissue variations for idMS/MS of M4. Node size is proportional to the cross-tissue relative intensity of each idMS/MS. Color mapping denotes rules presented in **(a)**. Grey nodes do not exhibit tissue specificity, whereas yellow nodes

were detected as tissue-specific. Red-circled nodes are annotated as flavonoid glycosides. ANT, anthers; COR, corolla tube; LIM, corolla limb; STY, style.

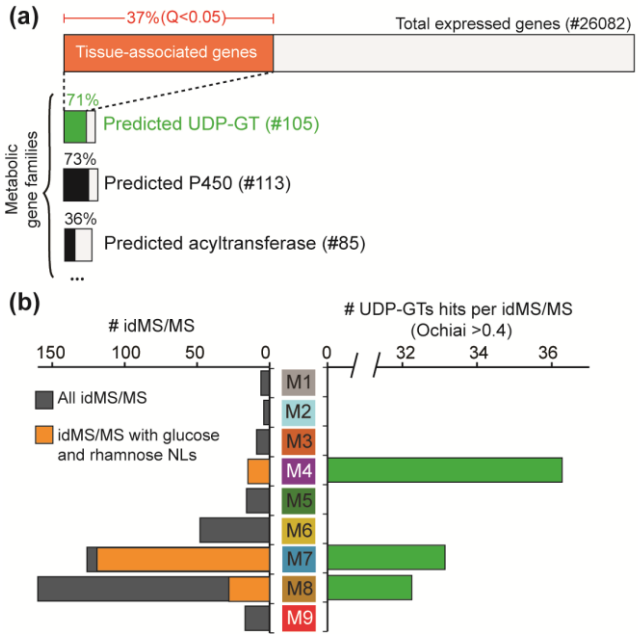

Figure 7. Prediction of tissue co-associations among UDP-glycosyltransferases (UDP-GT) and glycosylated metabolites. (a) Results of the Kurtosis filtering analysis for preferential tissue-gene associations. Examples are provided for the tissue-specificity of members of large metabolic gene families. Notably, 71% of all predicted UDP-glycosyltransferases exhibit tissue-specificity in the transcriptome data-set. **(b)** Left panel, number of tissue-specific idM/MS spectra containing glucose and rhamnose neutral losses (NLs) and therefore predicted as glycosylated secondary metabolites compared to the total number of tissue-specific idMS/MS per bi-clustering modules. M4 is enriched in flavonoid glycosides, M7 in 17-hydroxygeranyllinalool diterpene glycosides and M8 in *O*-acyl sugars. Right panel, the number of UDP-GT co-associated across tissues (Ochiai score > 0.4) with at least one idMS/MS containing glucose or rhamnose NL of each module.

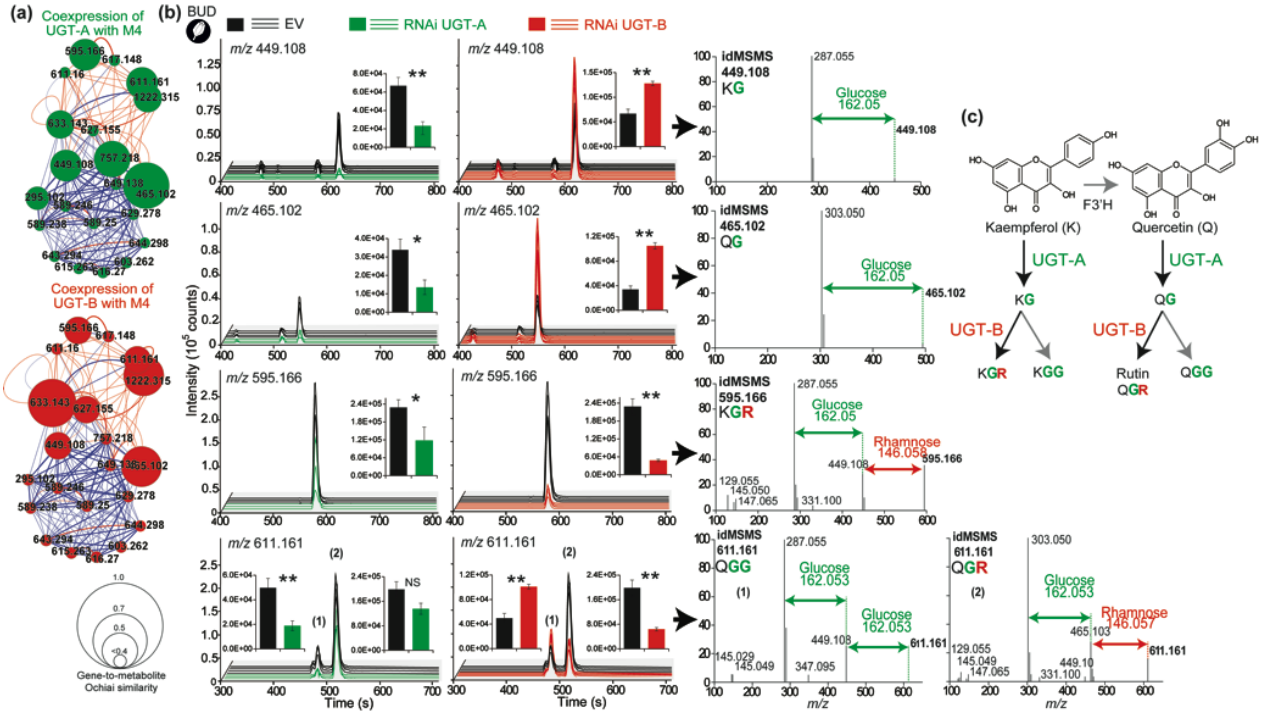

Figure 8. Silencing *UGT-A* and *UGT-B* reveals their involvement as UDP-glucosyl and rhamnosyltransferases, respectively, in floral flavonoid glycoside metabolism, two predictions of the tissue co-expression analysis. (a) Tissue co-association between *UDP-glycosyltransferase-A* (*UGT-A)* and *UGT-B* with idMS/MS of module M4. Size of network nodes is proportional to the strength of their co-associations (Ochiai score > 0.4). **(b)** UPLC-MS analysis of flower buds of plants inoculated with empty vector and gene silencing constructs for *UGT-A* and *UGT-B* (**Figure S19**). Chromatograms are traces corresponding to idMS/MS signals with strong co-association with these two UGTs. As supported by the annotation of idMS/MS spectra, silencing UGT-A decreases the glucosylation of flavonols while silencing *UGT-B* decreases their additional rhamnosylation. **(c)** Proposed metabolic scheme. Additional details on metabolite structures and on the pathway overview are provided in **Figure S20**. Asterisks denote significant differences between empty vector (EV) and UGT silenced lines (t-test, **P* < 0.05, ** < 0.01).

MATERIAL AND METHODS

Tissue-level metabolite extraction

Here we extracted 14 different tissues from 28 and 50-day-old *N. attenuata* plants growing in the glasshouse (Figure 1a). For non-reproductive tissues, the sample collection included: a pool of all non-senescing rosette leaves, combined lower, middle and higher segments of the stem, the complete root system and matured seeds. Reproductive parts were harvested as follows. Complete floral buds of 8 mm length – a stage at which the corolla has not yet protruded from the sepals and for which important gene expression and metabolic reconfigurations have been detected in previous work[48] – were harvested. Mature flowers at anthesis (5 days post-8mm stage, 7pm), were carefully separated into the following parts: the pedicel, complete sepal ring, the nectary, the ovary (not including the nectary), the style, anthers, filaments (not including anthers), the corolla tube (not including the limb) and the corolla limb.

Pools of 100 mg of isolated tissues (see Supplemental Methods) were extracted as follows using extraction buffers containing either 20% or 80% of methanol to increase the coverage of chemically-diverse metabolite classes. 1 mL extraction buffer (50 mM acetate buffer, pH 4.8, containing 20% or 80 % methanol) per 100 mg tissue was added and samples were homogenized in a ball mill (Genogrinder 2000; SPEX CertiPrep) for 45 s at 1× rate and 250 strokes per min. Homogenized samples were centrifuged at 16,000g, 4°C for 30 min, and supernatants were transferred into 1.5 mL microcentrifuge tubes and re-centrifuged as before. Supernatants of 400 µL were transferred to 2 mL glass vials for mass spectrometry-based metabolomics. In order to prevent the discarding of tissue-specific metabolites from the XCMS analysis due to poor grouping across samples **(see Supplemental Methods)**, 5 mixed extracts containing all 14 tissues at different ratios were generated and processed simultaneously with all other tissue samples.

Conditions for UPLC-ESI/TOF-MS profile mode analysis

An Acclaim column (150×2.1 mm, particle size 2.2 µm) with a 4 mm×4 mm guard column of the same material was used for the analysis. The following binary gradient was used with a Dionex Ultimate 3000 UPLC system: 0 to 1 min, isocratic 90% A (de-ionized water, 0.1% [v/v] acetonitrile and 0.05% formic acid), 10% B (acetonitrile and 0.05% formic acid); 1 to 40 min, gradient phase to 10% A, 85% B; 40 to 45 min, isocratic 10% A, 85% B. Flow rate was 300 µL/min. Eluted compounds were detected by a high-resolution MicroToF mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an electrospray ionization source operating in positive ionization mode. Typical instrument settings were as follows: capillary voltage 4500 V, capillary exit 130 V, dry gas temperature 200°C, dry gas flow of 8 L/min. Ions

were detected from *m/z* 50 to 1400 at a repetition rate of 1 Hz. Mass calibration was performed using sodium formate clusters (10 mM solution of NaOH in 50/50% v/v isopropanol/water containing 0.2% formic acid). Raw data files were converted to netCDF format using the export function of the Data Analysis v4.0 software (Bruker Daltonics, Bremen, Germany).

UHPLC-ESI/qTOF-MS conditions for indiscriminant MS/MS data acquisition

Data-independent or indiscriminant MS/MS fragmentation analysis (hereafter referred to as idMS/MS) was conducted in order to gain structural information on the overall metabolic profile detected by UHPLC-TOFMS (**see Supplemental Methods**). The UPLC binary gradient conditions previously used for the initial high throughput analysis in profile mode was again used for the idMS/MS analysis. The indiscriminant MS/MS data-set has been deposited in the open metabolomics database Metabolights (http://www.ebi.ac.uk) under the accession **MTBLS335**.

Assembly of compound-specific idMS/MS.

We used a previously designed precursor-to-product assignment pipeline[15] using the output results from processing with the R packages, XCMS and CAMERA. idMS/MS assembly was achieved via correlational analysis between MS1 and idMS/MS mass signals for low and high collision energies and newly implemented rules (**see Supplemental Methods**). The correlation analysis for precursor-to-product assignment was implemented using an R script and rules were operated using a C# script.

idMS/MS similarity scoring

idMS/MS were aligned in a pairwise manner and their similarity calculated according to two scores. First, a standard normalized dot product (NDP), also referred to as cosine correlation method, was used to score fragment similarity among spectra using the following equation:

$$NDP = \frac{\left(\sum_i^{S1\&S2} W_{S1,i}\, W_{S2,i}\right)^2}{\sum_i W_{S1,i}^2 \sum_i W_{S2,i}^2}$$

where S1 and S2 correspond respectively to spectrum 1 and spectrum 2 and $W_{S1,i}$ and $W_{S2,i}$ indicate peak intensity-based weights given to *i*th common peaks differing by less than 0.01Da between the two spectra. Weights were calculated as follows:

$$W = [Peak\ intensity]^m [Mass]^n$$

with m = 0.5 and n = 2 as suggested by MassBank.

A second scoring method involving the analysis of shared neutral losses among individual idMS/MS was implemented as described in the Supplemental Materials and Methods. For this, we used a list of 52 neutral losses (NLs) commonly encountered during tandem MS fragmentation (**Table S1**) as well as more specific ones that had been previously annotated for MS/MS spectra of *N. attenuata* secondary metabolite classes.

idMS/MS tissue-specificity inference using Kurtosis filtering

We used an outlier-insensitive Z-score measure -- generally considered preferable for the statistical description of sample groups containing extreme differences in values -- by using median and median absolute deviation (MAD) instead of mean and standard deviation for the normalization of both idMS/MS and RNAseq datasets in order to obtain relative expressions within tissues, as calculated using the following equation described in (Birmingham et al., 2009):

$$Z_i = \big(E_i - Median(E)\big)/MAD(E)$$

where E_i is the expression level of a metabolite or a gene in tissue *i*. *E* is a vector of a metabolite or a gene in all tissue samples.

Kurtosis was calculated for each metabolite and gene using an R package (moments) using the following equation:

$$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2}$$

where X_i stands for the expression level of a metabolite or a gene in *i*th tissue, $\bar{X}$ is the mean of the same metabolite or gene. P-value of the kurtosis was calculated using Anscombe.test function in the R moments package.

Tissue-specificity for a metabolite or a gene was defined using the reduction of kurtosis method as previously described[21]. When a leptokurtic expressed metabolite or gene removed high expression values for certain tissues, the kurtosis of the metabolite or the gene will be reduced. Threshold Z filtering of the data from a particular tissue was obtained by plotting the cumulative reductions in the kurtosis curves for any given kurtosis threshold using different Z threshold values (**File S7**). When defining FDR-adjusted P-value as Q, we chose a Z threshold of 2 for metabolite datasets where 99% of the metabolites with Q<0.01 exhibit

reduced kurtosis after applying the threshold cut-off. Similarly, a threshold of 3 was applied for the RNAseq dataset.

Gene-to-metabolite tissue-association similarities

Cross-tissue gene-to-metabolite associations provide valuable clues in formulating functional hypothesis about metabolic genes. To do so, we used as data input the idMS/MS and RNAseq binary data-sets computed separately for the following Z-scores: a Z-score of 1 indicating tissue-specificity and 0 indicating no tissue-specificity for a given feature. Similarities in gene and metabolite tissue-specificity were calculated using Ochiai coefficient calculated as following:

$$Och = \frac{a}{\sqrt{(a+b)}\sqrt{(a+c)}}$$

where a is the number of tissue-associations for which both a metabolite and a gene exhibit a Z-score of 1, b is the number of tissue-associations where a metabolite is 1 and the gene is 0, c is the number of tissue-associations where the metabolite is 0 and the gene is 1.

idMS/MS molecular networking by bi-clustering

To perform this clustering, we used the R package DiffCoEx which is based an extension of the Weighted Gene Coexpression Analysis (WGCNA). Using NDP and NL-scoring matrices for 895 idMS/MS spectra, we computed a comparative correlation matrix using DiffCoEx with the parameters of “cutreeDynamic” set to method="hybrid", cutHeight = 0.999, deepSplit = T, minClusterSize = 10. The R source code of DiffCoEx is downloaded from additional file 1 in Tesson et al. (2010)[49], the required R WGCNA package can be found at http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/Rpackages/WGCNA.

Virus-induced gene silencing (VIGS)

Vector construction, plant growth, and inoculation conditions were as described by Saedler and Baldwin (2004)[50]. Briefly, 200- to 300-bp fragments of *N. attenuata* target genes were amplified by PCR using specific primer pairs as listed in **Table S2**. Amplified fragments were cloned into pTV00 vector, and plasmids were transformed by electroporation into *Agrobacterium tumefaciens* strain GV3101. A pTV00 plasmid without insert (EV) was used as a negative control in all experiments. Three leaves of 24- to 25-d-old *N. attenuata* plants were infiltrated with a 1:1 mixture of *A. tumefaciens* transformed with pBINTRA and one of the gene fragment-containing construct or the pTV00 construct. *Phytoene desaturase* (pTVPDS) causing bleaching of tobacco leaves due to the depletion of carotenoids was used as a positive control to monitor the progression of VIGS in a separate set of inoculated plants.

VIGS-silenced plants were used for treatment after PDS-VIGS leaves developed a strong bleaching phenotype. Silencing efficiency was verified by RT-qPCR of target gene transcripts after RNA extraction and cDNA synthesis.

RT-qPCR analysis of gene silencing efficiency

Total RNA was extracted by adding Trizol reagent (Invitrogen; http://www.invitrogen.com) to approximately 150 mg of powdered leaf material ground in liquid nitrogen following the manufacturer's protocol. A total of 500 ng of DNA-free RNA samples was reverse transcribed using oligo(dT)18 primers and SuperScript II enzyme (Invitrogen) following the manufacturer's recommendations. All RT-qPCR assays were performed with a Stratagene MX3005P instrument (http://www.stratagene.com) as recommended by the manufacturer. To normalize transcript levels, primers specific for the elongation factor-1α gene from *Nicotiana tabacum* (EF1-α; accession no. D63396) were used. Specific primers in the 5′ to 3′ direction used for SYBR Green-based analyses are listed in **Table S2**.

RNAseq dataset of different tissues and data mining

A list of the tissues collected for RNAseq analysis and of meta-data related to this experiment is available in **Figure S14**. This experiment had been conducted and involved additional tissues and physiological conditions in addition to those reported in the metabolomics study presented here.

ACKNOWLEDGMENTS

We thank Dr. Mathias Schöttner for technical support in establishing of idMS/MS acquisition method and Dr. Klaus Gase for help with gene silencing construct design. DL and ITB are funded by the Max-Planck-Society, an Advanced Grant no. 293926 of the European Research Council to ITB and by the Collaborative Research Centre "Chemical Mediators in Complex Biosystems - ChemBioSys" (SFB 1127). E.G.'s research in Heidelberg is supported within the framework of the Deutsche Forschungsgemeinschaft Excellence Initiative to the University of Heidelberg.

REFERENCES

1 Weng, J. K., Philippe, R. N. & Noel, J. P. The rise of chemodiversity in plants. *Science* **336**, 1667-1670 (2012).

2 Wink, M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**, 3-19 (2003).

3 Wink, M. & Carey, D. B. Variability of quinolizidine alkaloid profiles of *Lupinus*

argenteus (Fabaceae) from North-America. *Biochem Syst Ecol* **22**, 663-669 (1994).

4 Itkin, M. *et al.* GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. *The Plant cell* **23**, 4507-4525 (2011).

5 vonPoser, G. L., Toffoli, M. E., Sobral, M. & Henriques, A. T. Iridoid glucosides substitution patterns in Verbenaceae and their taxonomic implication. *Plant Syst Evol* **205**, 265-287 (1997).

6 Matsuda, F. *et al.* AtMetExpress development: A phytochemical atlas of *Arabidopsis* development. *Plant physiology* **152**, 566-578 (2010).

7 Tissier, A. Glandular trichomes: what comes after expressed sequence tags? *Plant J* **70**, 51-68 (2012).

8 Schilmiller, A. L. *et al.* Studies of a biochemical factory: tomato trichome deep expressed sequence tag sequencing and proteomics. *Plant physiology* **153**, 1212-1223 (2010).

9 Zang, Y. X. *et al.* Genome-wide identification of glucosinolate synthesis genes in *Brassica rapa*. *Febs J* **276**, 3559-3574 (2009).

10 Hirai, M. Y. *et al.* Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *P Natl Acad Sci USA* **104**, 6478-6483 (2007).

11 Rajniak, J., Barco, B., Clay, N. K. & Sattely, E. S. A new cyanogenic metabolite in *Arabidopsis* required for inducible pathogen defence. *Nature* **525**, 376-+ (2015).

12 Sakurai, T. *et al.* PRIMe update: Innovative content for plant metabolomics and integration of gene expression and metabolite accumulation. *Plant Cell Physiol* **54**, E5-+ (2013).

13 Bowen, B. P. & Northen, T. R. Dealing with the unknown: metabolomics and metabolite atlases. *J Am Soc Mass Spectr* **21**, 1471-1476 (2010).

14 Allard, P. M. *et al.* Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. *Anal Chem* **88**, 3317-3323 (2016).

15 Broeckling, C. D., Heuberger, A. L., Prince, J. A., Ingelsson, E. & Prenni, J. E. Assigning precursor-product ion relationships in indiscriminant MS/MS data from non-targeted metabolite profiling studies. *Metabolomics* **9**, 33-43 (2013).

16 Li, D. P., Baldwin, I. T. & Gaquerel, E. Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. *P Natl Acad Sci USA* **112**, E4147-E4155 (2015).

17 Gaquerel, E., Kuhl, C. & Neumann, S. Computational annotation of plant metabolomics profiles via a novel network-assisted approach. *Metabolomics* **9**, 904-918 (2013).

18 Shannon, C. E. A mathematical theory of communication. *At&T Tech J* **27**, 379-423 (1948).

19 Martinez, O. & Reyes-Valdes, M. H. Defining diversity, specialization, and gene specificity in transcriptomes through information theory. *P Natl Acad Sci USA* **105**, 9709-9714 (2008).

20 Heiling, S. *et al.* Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *The Plant cell* **22**, 273-292, doi:10.1105/tpc.109.071449 (2010).

21 Li, S. *et al.* Gene-sharing networks reveal organizing principles of transcriptomes in *Arabidopsis* and other multicellular organisms. *The Plant cell* **24**, 1362-1378 (2012).

22 Watrous, J. *et al.* Mass spectral molecular networking of living microbial colonies. *P Natl Acad Sci USA* **109**, E1743-E1752 (2012).

23 Nomura, T., Murase, T., Ogita, S. & Kato, Y. Molecular identification of tuliposide B-converting enzyme: a lactone-forming carboxylesterase from the pollen of tulip. *Plant J* **83**, 252-262 (2015).
24 Ulanowicz, R. E. Information theory in ecology. *Comput Chem* **25**, 393-399 (2001).
25 Eren, A. M., Borisy, G. G., Huse, S. M. & Welch, J. L. M. Oligotyping analysis of the human oral microbiome. *P Natl Acad Sci USA* **111**, E2875-E2884 (2014).
26 Moco, S. *et al.* Tissue specialization at the metabolite level is perceived during the development of tomato fruit. *J Exp Bot* **58**, 4131-4146 (2007).
27 Kessler, D., Diezel, C. & Baldwin, I. T. Changing pollinators as a means of escaping herbivores. *Curr Biol* **20**, 237-242 (2010).
28 Euler, M. & Baldwin, I. T. The chemistry of defense and apparency in the corollas of *Nicotiana attenuata*. *Oecologia* **107**, 102-112 (1996).
29 Kessler, D. & Baldwin, I. T. Making sense of nectar scents: the effects of nectar secondary metabolites on floral visitors of *Nicotiana attenuata*. *Plant J* **49**, 840-854 (2007).
30 Baldwin, I. T., Staszakkozinski, L. & Davidson, R. Up in smoke: I. Smoke-derived germination cues for postfire annual, *Nicotiana attenuata* torr. Ex. Watson. *Journal of chemical ecology* **20**, 2345-2371 (1994).
31 Baldwin, I. T. & Morse, L. Up in smoke: II. Germination of *Nicotiana attenuata* in response to smoke-derived cues and nutrients in burned and unburned soils. *Journal of chemical ecology* **20**, 2373-2391 (1994).
32 Baldwin, I. T. An ecologically motivated analysis of plant-herbivore interactions in native tobacco. *Plant physiology* **127**, 1449-1458 (2001).
33 Gulati, J., Kim, S. G., Baldwin, I. T. & Gaquerel, E. Deciphering herbivory-induced gene-to-metabolite dynamics in *Nicotiana attenuata* tissues using a multifactorial approach. *Plant physiology* **162**, 1042-1059 (2013).
34 Weinhold, A. & Baldwin, I. T. Trichome-derived *O*-acyl sugars are a first meal for caterpillars that tags them for predation. *P Natl Acad Sci USA* **108**, 7855-7859 (2011).
35 Onkokesung, N. *et al.* MYB8 controls inducible phenolamide levels by activating three novel hydroxycinnamoyl-coenzyme A:polyamine transferases in *Nicotiana attenuata*. *Plant physiology* **158**, 389-407 (2012).
36 Mckey, D. Adaptive patterns in alkaloid physiology. *Am Nat* **108**, 305-320 (1974).
37 McKey, D. The distribution of secondary compounds within plants. *Herbivores: their interaction with secondary plant metabolites. Academic Press, New York* 55-133 (1979).
38 Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science* **347** (2015).
39 Schmid, M. *et al.* A gene expression map of *Arabidopsis thaliana* development. *Nature genetics* **37**, 501-506 (2005).
40 Bassard, J. E., Ullmann, P., Bernier, F. & Werck-Reichhart, D. Phenolamides: Bridging polyamines to the phenolic metabolism. *Phytochemistry* **71**, 1808-1824 (2010).
41 Werner, C., Hu, W. Q., Lorenziriatsch, A. & Hesse, M. Di-coumaroylspermidines and tri-coumaroylspermidines in anthers of different species of the genus *Aphelandra*. *Phytochemistry* **40**, 461-465 (1995).
42 Meurer, B., Wiermann, R. & Strack, D. Phenylpropanoid patterns in *Fagales* pollen and their phylogenetic relevance. *Phytochemistry* **27**, 823-828 (1988).
43 Matsuno, M. *et al.* Evolution of a novel phenolic pathway for pollen development. *Science* **325**, 1688-1692 (2009).
44 Yang, R. L. & Wang, X. F. Organ evolution in angiosperms driven by correlated divergences of gene sequences and expression patterns. *The Plant cell* **25**, 71-82

(2013).
45 Yonekura-Sakakibara, K. *et al.* Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. *The Plant cell* **20**, 2160-2176 (2008).
46 Tohge, T. & Fernie, A. R. Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nat Protoc* **5**, 1210-1227 (2010).
47 Ma, S. M. *et al.* Organization of the mammalian metabolome according to organ function, lineage specialization, and longevity. *Cell Metab* **22**, 332-343 (2015).
48 Stitz, M., Hartl, M., Baldwin, I. T. & Gaquerel, E. Jasmonoyl-l-isoleucine coordinates metabolic networks required for anthesis and floral attractant emission in wild tobacco (*Nicotiana attenuata*). *The Plant cell* **26**, 3964-3983 (2014).
49 Tesson, B. M., Breitling, R. & Jansen, R. C. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC bioinformatics* **11** (2010).
50 Saedler, R. & Baldwin, I. T. Virus-induced gene silencing of jasmonate-induced direct defences, nicotine and trypsin proteinase-inhibitors in *Nicotiana attenuata*. *J Exp Bot* **55**, 151-157 (2004).

SUPPLEMENTAL MATERIAL

Supplemental Figures

Figure S1. Principal component analysis (PCA) of the UHPLC-ToFMS metabolic profiles for all tissue and extraction types.

Figure S2. Examples of reconstructed idMS/MS for representatives of the main secondary metabolite classes.

Figure S3. Heatmaps for idMS/MS relative expression across tissues and extraction conditions.

Figure S4. Information theory-based analysis of the degree of specialization and diversity in the idMS/MS composition across tissues.

Figure S5. Tissue-based variations in the accumulation of HGL-DTGs.

Figure S6. Tissue-based variations in the accumulation of phenolic derivatives.

Figure S7. Significant levels of tissue-specificity derived from Kurtosis filtering for particular metabolic steps in the HGL-DTG pathway.

Figure S8. Significant levels tissue-specificity derived from Kurtosis filtering for several phenolic derivatives.

Figure S9. Implementing the reduction of Kurtosis analysis to infer idMS/MSs with high tissue-specificity.

Figure S10. Different degrees of idMS/MS tissue-specificity inferred from the Kurtosis reduction analysis.

Figure S11. Tissue-network based on idMS/MSs with preferential tissue expression.

Figure S12. Dot-plots depicting the tissue-based distribution idMS/MS classified by structural similarity into modules 1 to 9.

Figure S13. Implementing the reduction of Kurtosis analysis to infer gene with high tissue-specificity.

Figure S14. Gene-tissue specificity and gene-sharing among tissues of the transcriptome data-set.

Figure S15. GO enrichment analysis for the total gene pool exhibiting preferential tissue association and for those with significant co-tissue association with idMS/MS classified as parts of modules M4 and M8.

Figure S16. A sub-group of module 8 enriched in *O*-acyl sugar metabolites and GO terms from transcriptomic data sharing significant tissue co-associations with targeted metabolites.

Figure S17. A sub-group of module 4 enriched in flavonoids and GO terms from transcriptomic data sharing significant tissue co-associations with targeted metabolites.

Figure S18. Phylogenetic tree analysis for the UDP-glucosyltransferases tested during VIGS experiments.

Figure S19. Gene silencing efficiency for the UDP-glucosyltransferases tested during VIGS experiments.

Figure S20. Proposed biosynthetic scheme for main flavonoid glycosides detected in this study.

Supplemental Tables

Table S1. List of the 52 neutral losses used for neutral loss similarity calculation

Table S2. List of primers used for qRT-PCR and gene fragment cloning for VIGS

Supplemental Files

File S1. XCMS-processed data-set.

File S2. ".msp" file library of deconvoluted idMS/MS

File S3. Structural and tissue annotation of the idMS/MS biclustering and cross-tissue ZMAD scaled intensities in the two extraction systems.

Supplemental files can be found via dropbox link:
https://www.dropbox.com/sh/gdkgzr2fsoxu9zd/AAC5XrFQzjetq0KALzr5OWISa?dl=0

Figures

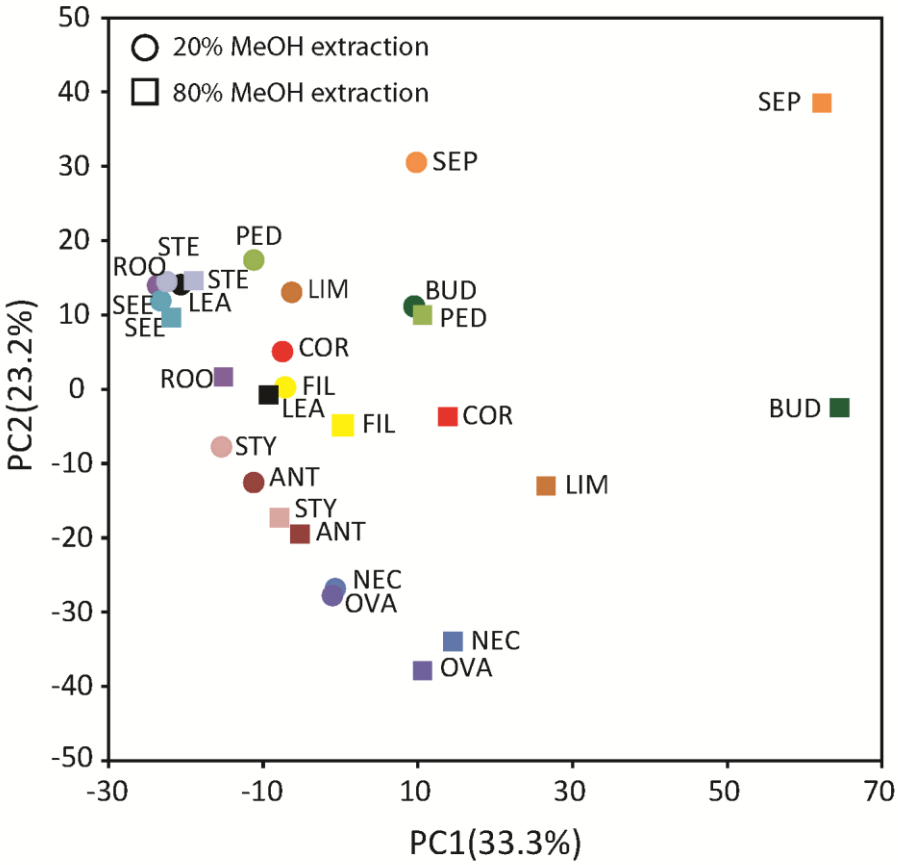

Figure S1. Principal component analysis (PCA) of the UHPLC-ToFMS metabolic profiles for all tissue and extraction types. The PCA score plot was conducted on the auto-scaled complete mass feature matrix resulting from XCMS processing of the samples extracted by 20% and 80% methanol which are represented as circle and rectangular shapes respectively. Detailed explanations of the tissue collection procedure are provided in the Method section.

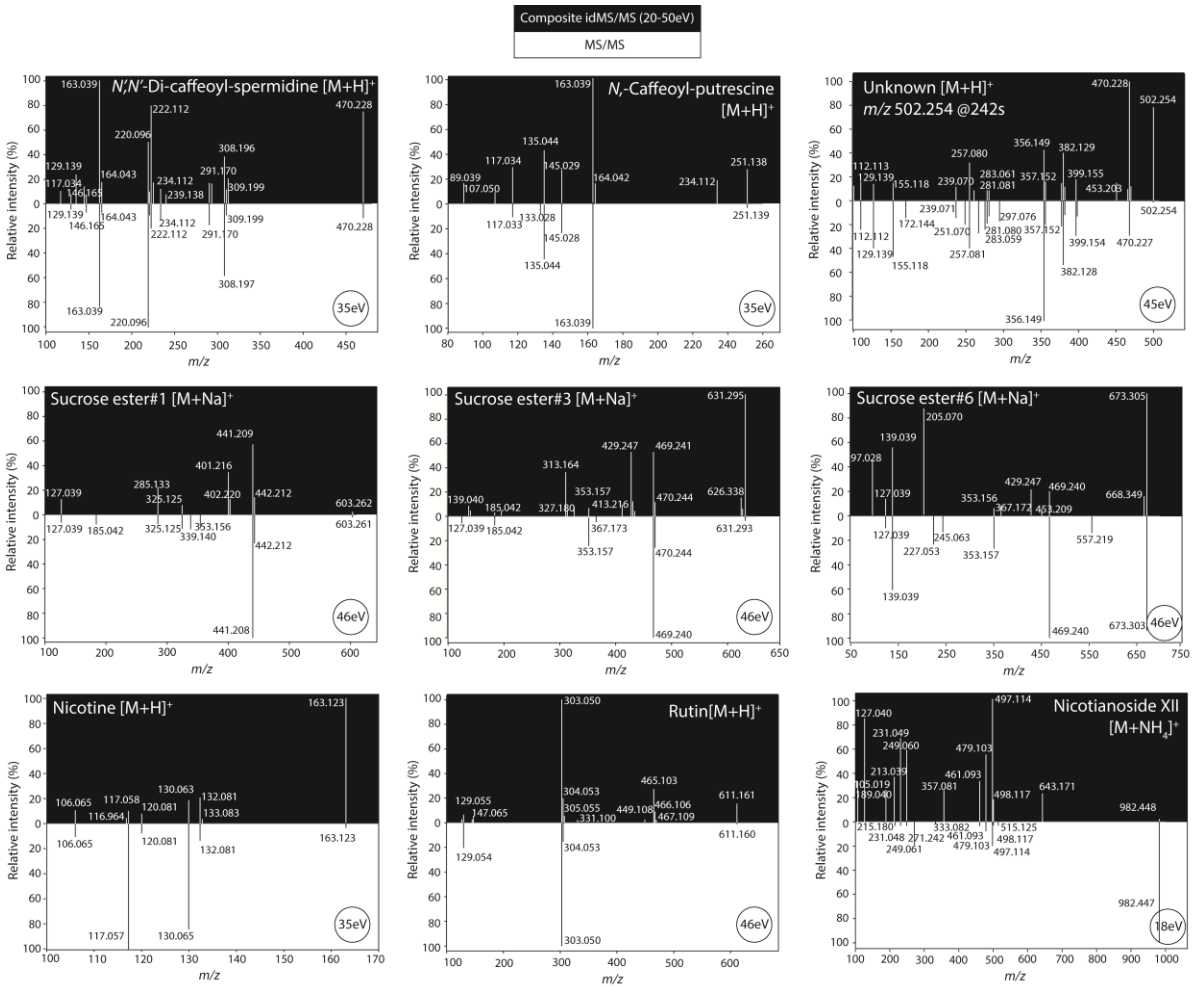

Figure S2. Examples of reconstructed idMS/MS spectra for representatives of main secondary metabolite classes. Spectra with a black background correspond to idMS/MS assembled in the context of this study for different collision energies (see Methods). Lower spectra were obtained by Gaquerel et al.[1], Heiling et al.[2] for the same precursor ions using an optimized collision energy value as shown in the circle at the bottom.

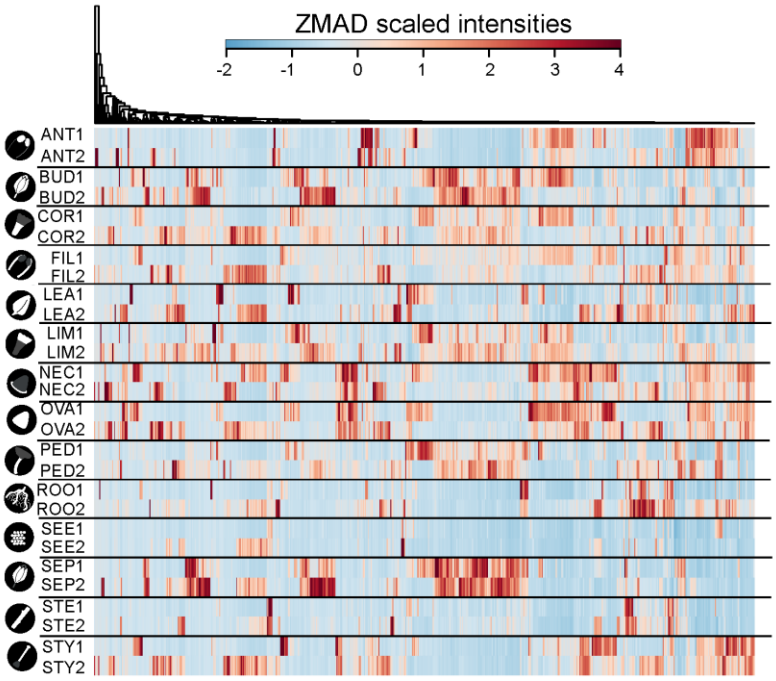

Figure S3. Heatmaps for idMS/MS relative expression across tissues and extraction conditions. Heatmap after hierarchical clustering (upper tree, Euclidean distance as metric) of idMS/MS relative intensities (Z-score of the median absolute distance [ZMAD], see Methods). Lines delimit heatmap vectors for each tissue types. 1 refers to an extraction with 20% methanol, 2 with 80% methanol. Tissue vectors are not hierarchically clustered. ZMAD expression data are reported in **File S1**. Detailed explanations of the tissue collection procedure are provided in the Method section.

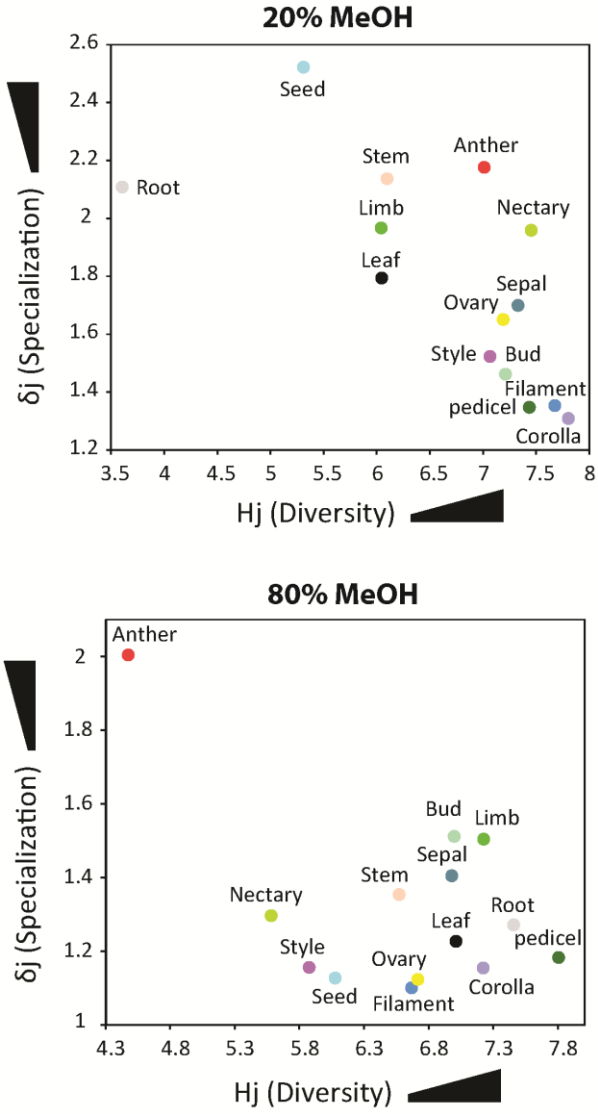

Figure S4. Information theory-based analysis of the degree of specialization and diversity in the idMS/MS composition across tissues. Tissue specialization (δj) and diversity (Hj) are mapped in a two-dimensional space using two indexes, where Hj, tissue-level idMS/MS diversity is calculated by Shannon entropy of idMS/MS frequency distribution of each tissue and δj, tissue-level idMS/MS specialization is measured by the average specificity of each idMS/MS component by taking into consideration its frequency among tissues. Tissue metabolomes extracted for two different extraction conditions (20% or 80% methanol) were separately analyzed and visualized into two panels and tissue types are differentiated by different colors.

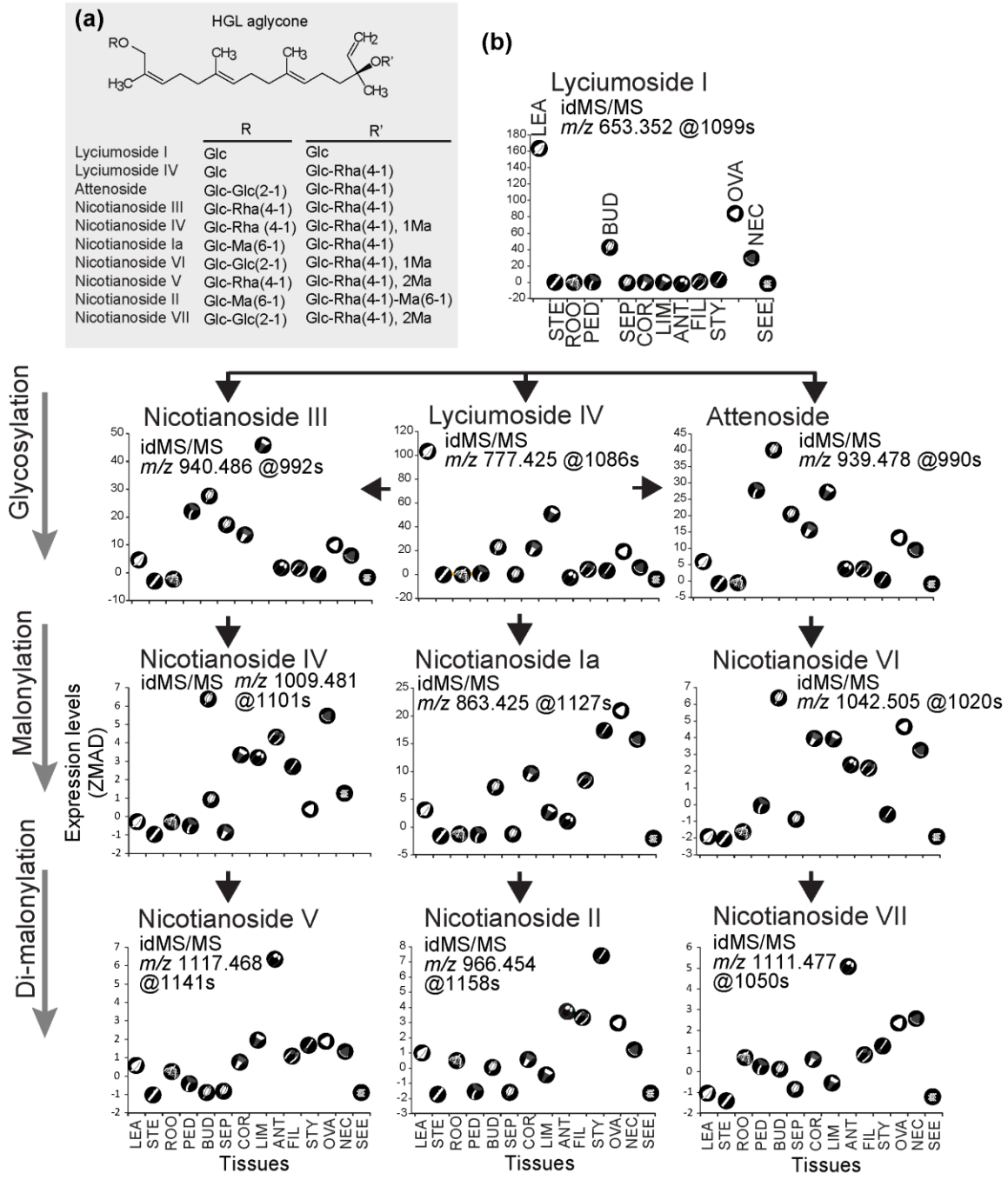

Figure S5. Tissue-based variations in the accumulation of HGL-DTGs. (a) 17-hydroxygeranyllinalool (HGL) diterpene glycosides are abundant secondary metabolites in *N. attenuata* which differ in the number and types of sugar (Glc, glucose; Rha, Rhamnose; with or without malonyl, Ma, groups) decorations added to the acyclic HGL backbone which is characteristic of this compound family. **(b)** Z-score-normalized median absolute distances are employed to visualize cross-tissue variations in the idMS/MSs corresponding to the main HGL-DTG intermediates.

Important changes in cross-tissue variations are detected along the HGL-DTG metabolic pathway and notably a progressive enrichment of certain metabolites in reproductive floral tissues.

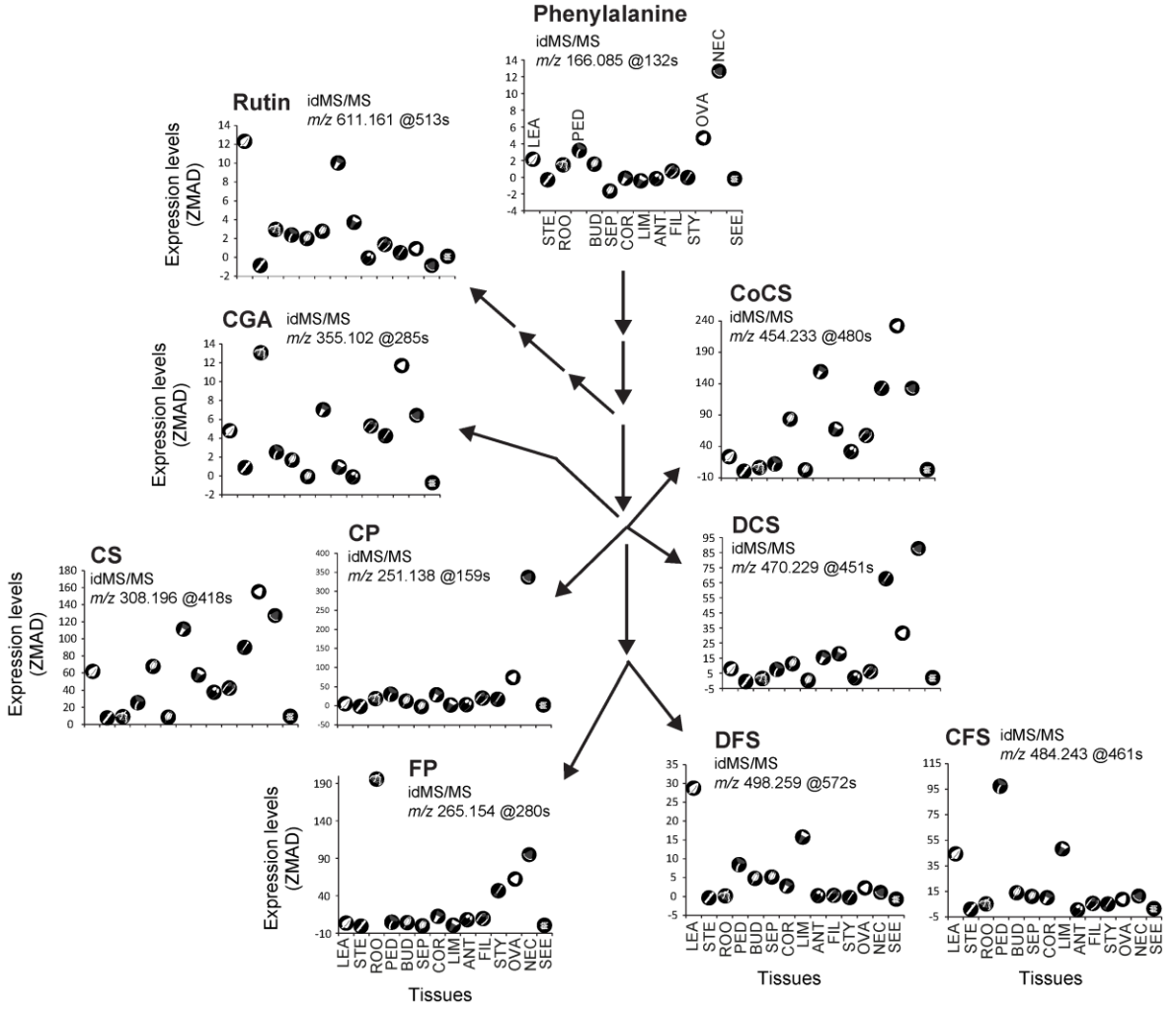

Figure S6. Tissue-level variations in the accumulation of phenolic derivatives. Schematic of the core phenylpropanoid pathway and major phenylpropanoid-quinate and -polyamine conjugates. Z-score-normalized median absolute distances are used to visualize cross-tissue variations in the idMS/MSs corresponding to metabolic intermediates. Important changes in cross-tissue variations are detected along branches of the metabolic pathway.

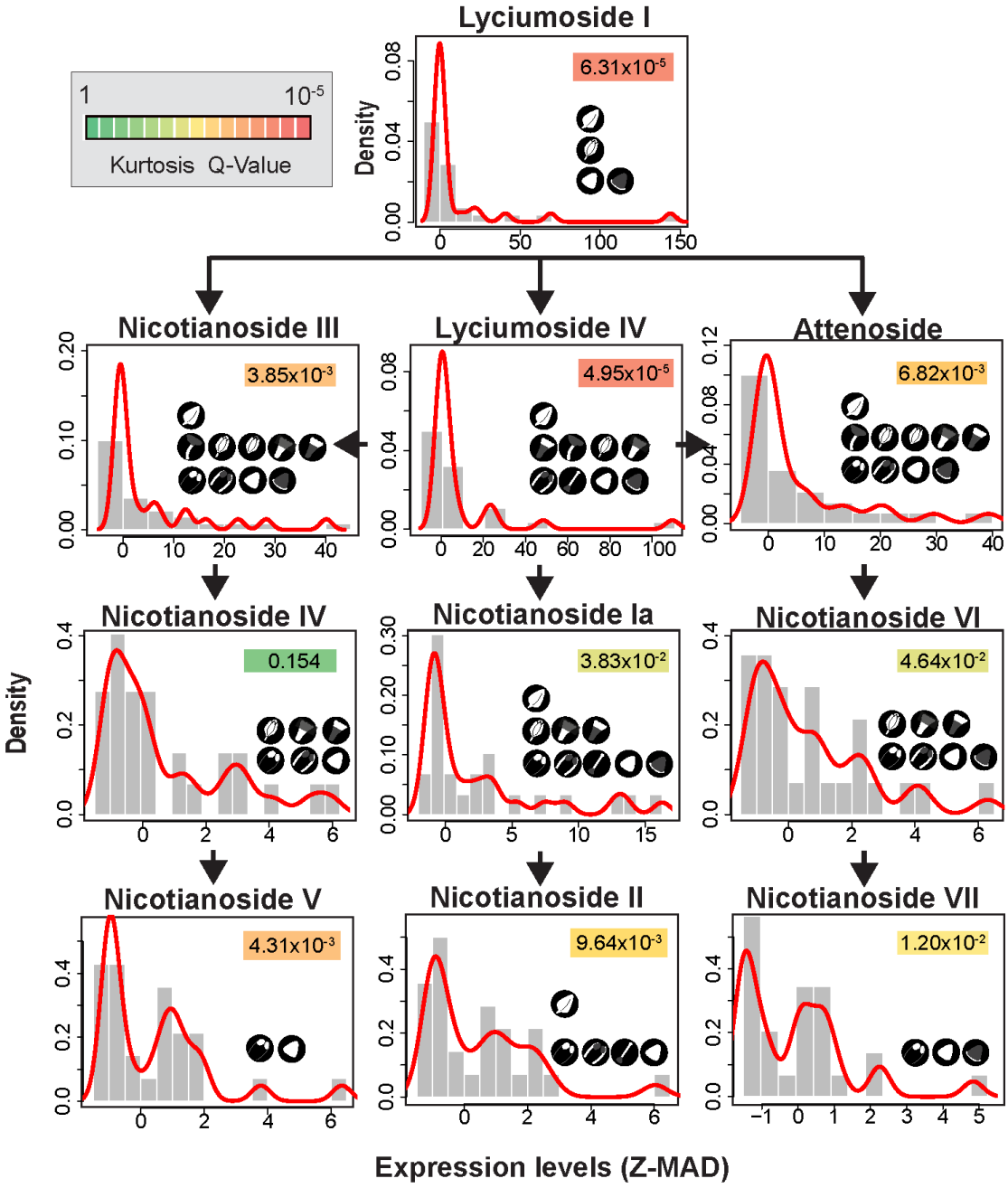

Figure S7. Significant levels of tissue-specificity derived from Kurtosis filtering for particular metabolic steps in the HGL-DTG pathway. Q-values present in color boxes derived from Kurtosis filtering analysis (see Methods) for the distribution of idMS/MSs corresponding to particular HGL-DTG (17-hydrogeranyllinalool diterpene glycosides). Associated tissues for each idMS/MS are presented as tissue icons below the Q-value box. Low Q-values indicate strong tissue-preferentiality in the

accumulation of particular HGL-DTGs. Strongest tissue-associations are detected for upstream steps in the pathway indicating clear tissue-specificities in the biosynthesis of these upstream intermediates from which the complete HGL-DTG chemotype derives from. Complete results of the Kurtosis analysis are presented in **File S1**.

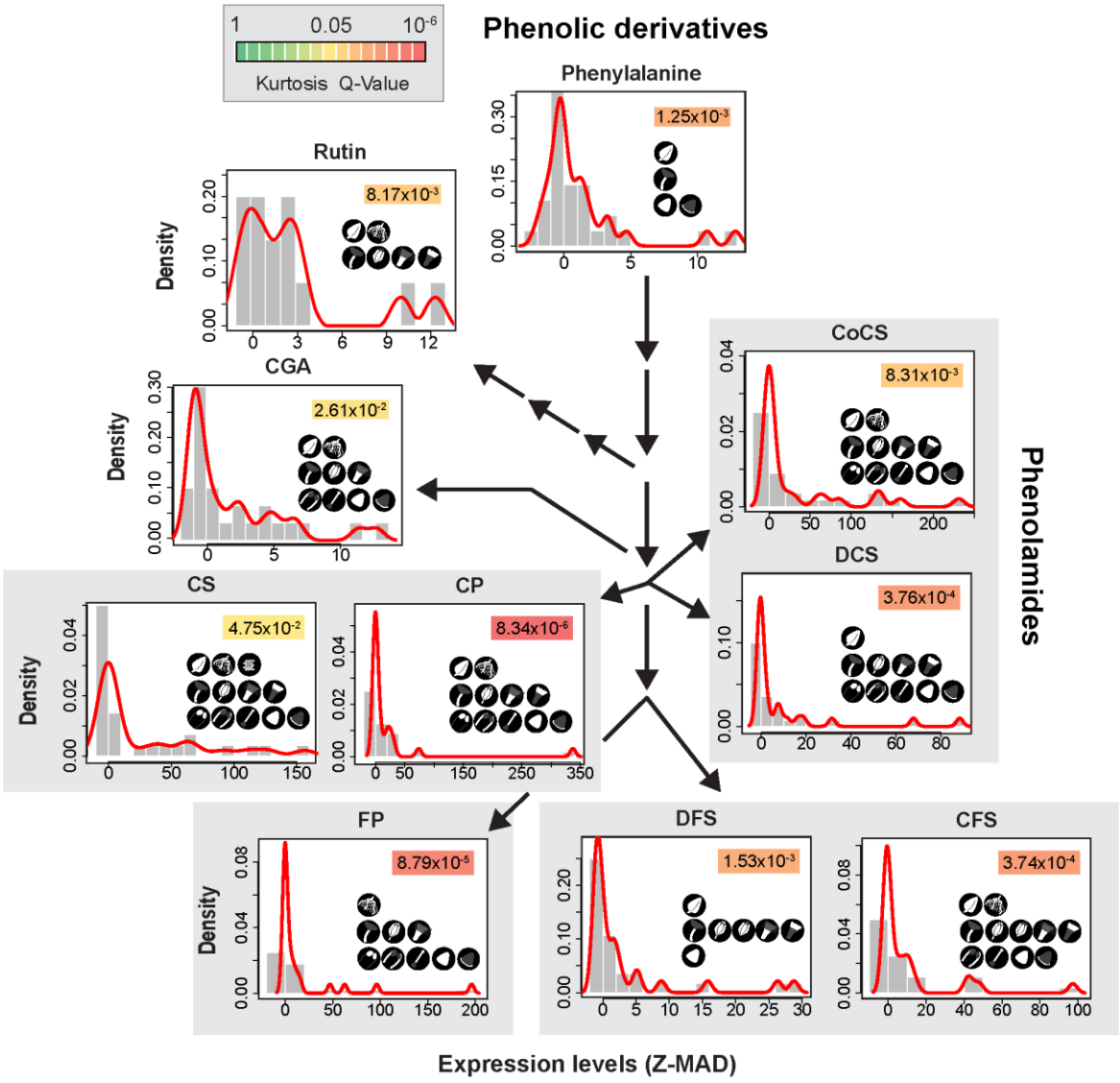

Figure S8. Significant levels tissue-specificity derived from Kurtosis filtering for several phenolic derivatives. Q-values present in colored boxes derived from Kurtosis filtering analysis (see Methods) for the distribution of idMS/MSs corresponding to particular phenolic derivatives. Associated tissues for each idMS/MS are presented as tissue icons below the Q-value box. Low Q-values indicate strong tissue-preferentiality in the accumulation of particular phenolic derivatives. Complete results of the Kurtosis analysis are presented in **File S1**. CFS, *N',N''*-caffeoyl,feruloyl-spermidine; CGA, Chlorogenic acid; CoCS, *N',N''*-coumaraoyl,caffeoyl-spermidine; CP, *N*-caffeoylputrescine; CS, *N*-caffeoylspermine; FP, *N*-feruloylputrescine; DCS, *N', N''*-dicaffeoylspermidine; DFS,

N',N''-caffeoyl, feruloyl-spermidine.

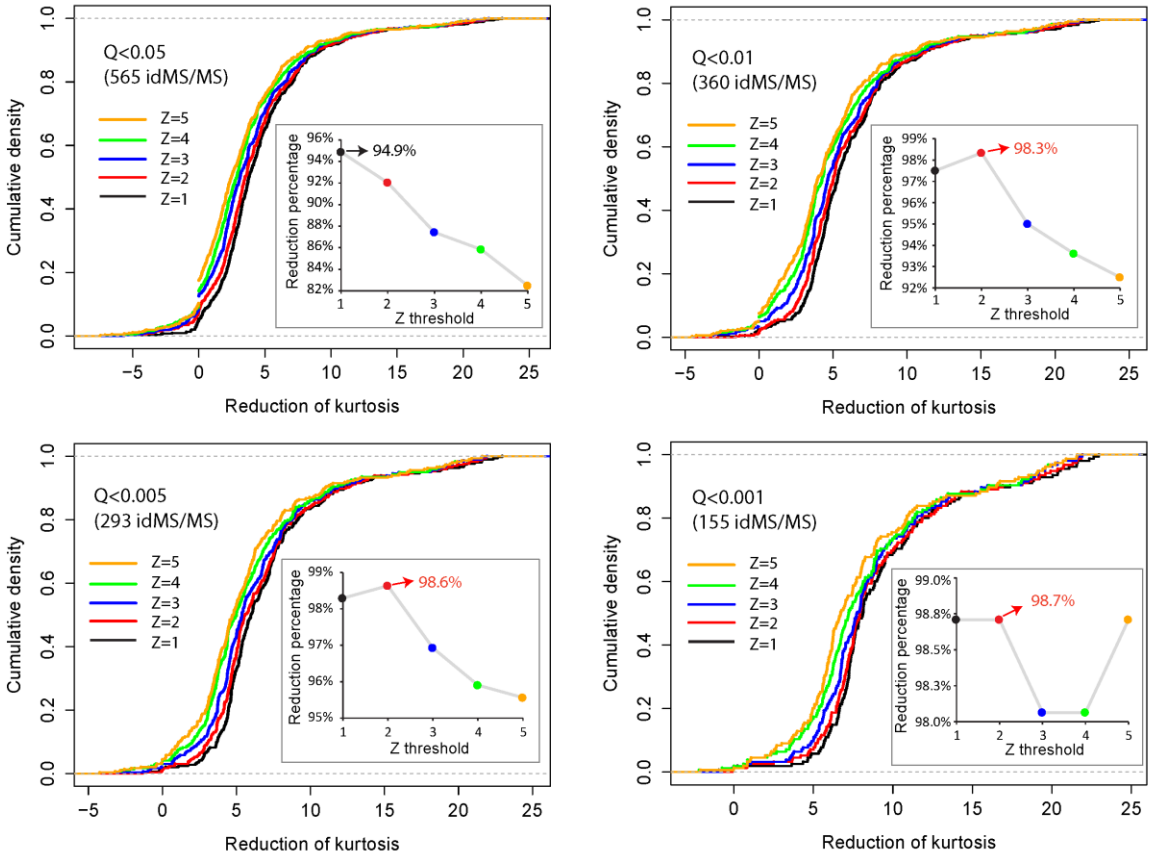

Figure S9. Implementing the reduction of Kurtosis analysis to infer idMS/MSs with high tissue-specificity. Cumulative distribution of idMS/MSs cross-tissue expression generated for different q-value (Q) thresholds obtained from R qvalue package[3]. The analysis is based on the Anscombe test for kurtosis using the Anscombe.test function in the R moments package as described in Li et al.[4] and in the methods section. The x axis reports on the Kurtosis reduction when a certain Z threshold (ZMAD-normalized expression values) was applied, the y axis reports on the corresponding cumulative density. The insert panels correspond to the Kurtosis reduction percentage of idMS/MSs when a given Z threshold was selected. Different colors denote for different Z threshold and the corresponding cumulative curves. Z=2 was selected as the threshold to extract idMS/MSs with leptokurtic behavior.

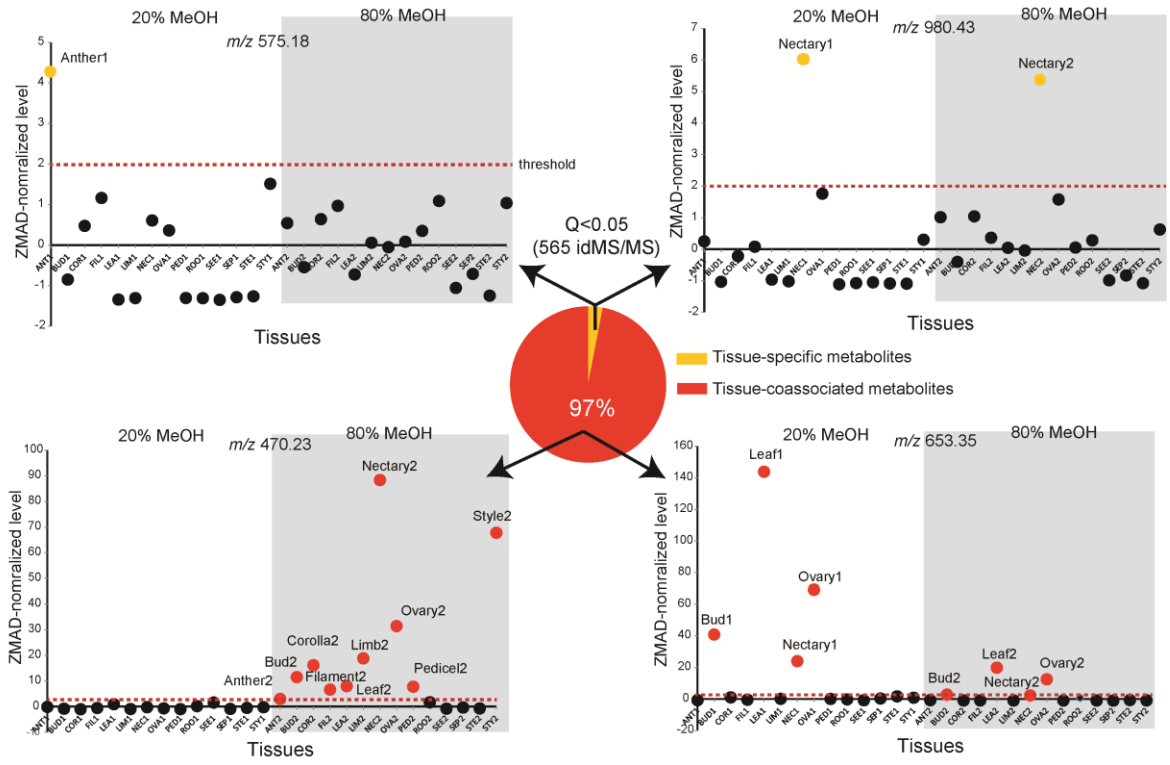

Figure S10. Different degrees of idMS/MS tissue-specificity inferred from the Kurtosis reduction analysis. The kurtosis analysis was used to discriminate leptokurtic idMS/MSs from those distributed in all tissues and the Z threshold of 2 that obtained from the reduction of Kurtosis analysis enabled the detection of single-tissue-specific and tissue-co-associated metabolites. The approach not only extracts single-tissue-specific idMS/MSs (3% of the 565 idMS/MSs with a Q<0.05 for the Kurtosis analysis; two upper panel as examples) but also those preferentially accumulating in several tissues (97% of idMS/MS; two lower panels as examples). 1 and 2 indices after tissue names refer to the 20 % and 80 % methanol respectively.

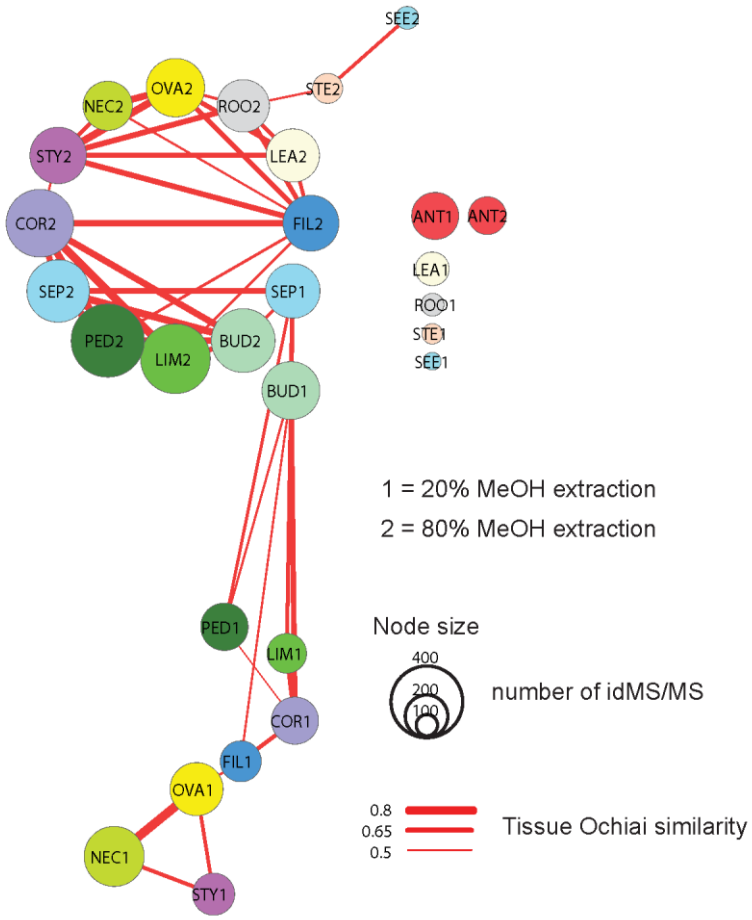

Figure S11. Tissue-network based on idMS/MSs with preferential tissue expression. The analysis is based on the number of shared tissue-specific idMS/MSs between tissue pairs inferred from the reduction of Kurtosis analysis of tissue-specificity. 1 and 2 indices refer to the type of extraction procedure and node size is proportional to the number of idMS/MSs per tissue type. The Ochiai coefficient (see methods) was used to measure similarity (number of shared idMS/MSs) between tissue types and is in proportion to edge thickness.

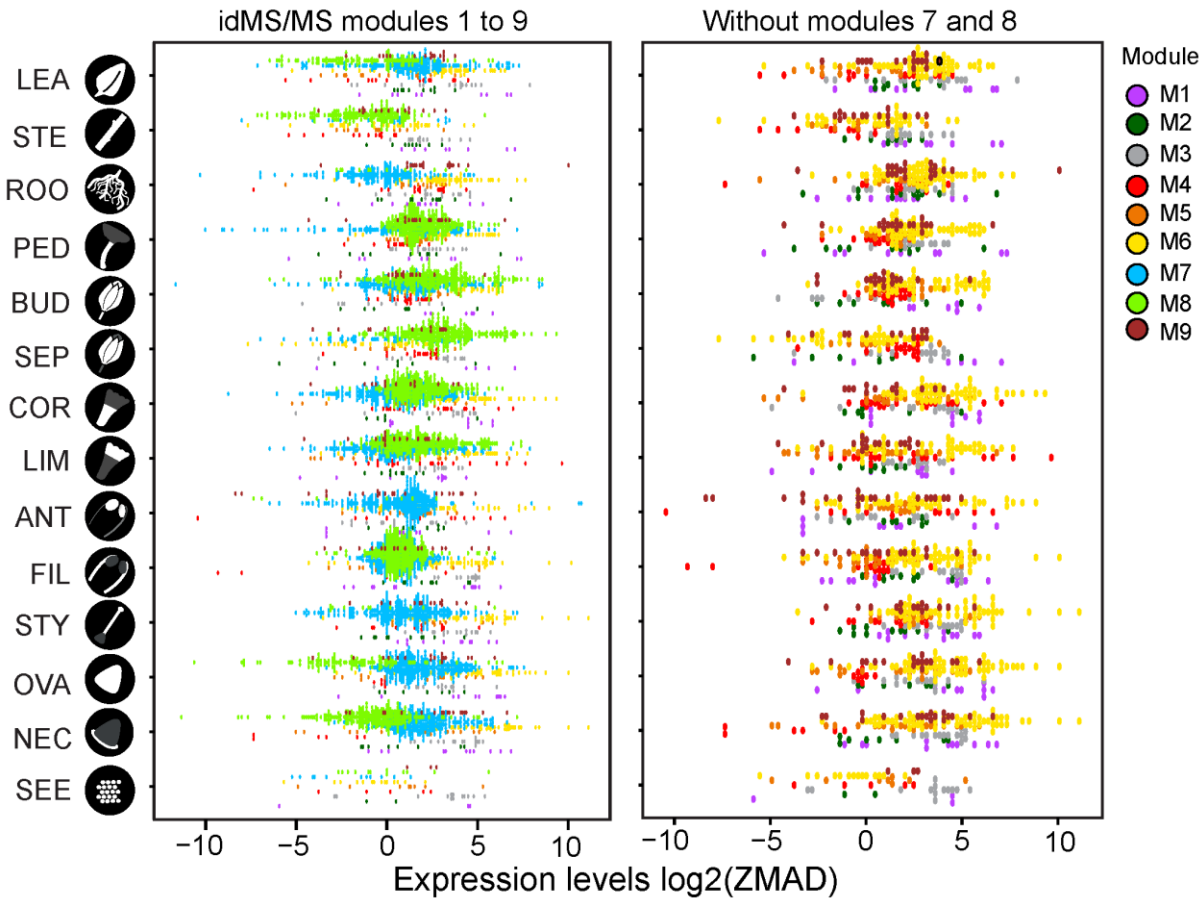

Figure S12. Dot-plots depicting the tissue-based distribution of idMS/MSs classified by structural similarity into modules 1 to 9. The 9 modules derive from the similarity classification of idMS/MS spectra according to Figure 5. Relative expression values across tissues as captured by the *Z*-score of the median absolute distance (ZMAD) were log2-transformed. Modules 7 (enriched in HGL-DTGs colored in blue) and 8 (enriched in *O*-acyl sugars colored in green) cluster the largest number of idMS/MSs. idMS/MSs from these two clusters poorly populate root, style and anther samples in the case of M8 and stem sample in the case of M7. A complete matrix reporting on idMS/MS module classification and ZMAD score in different tissues is available in **File S3**.

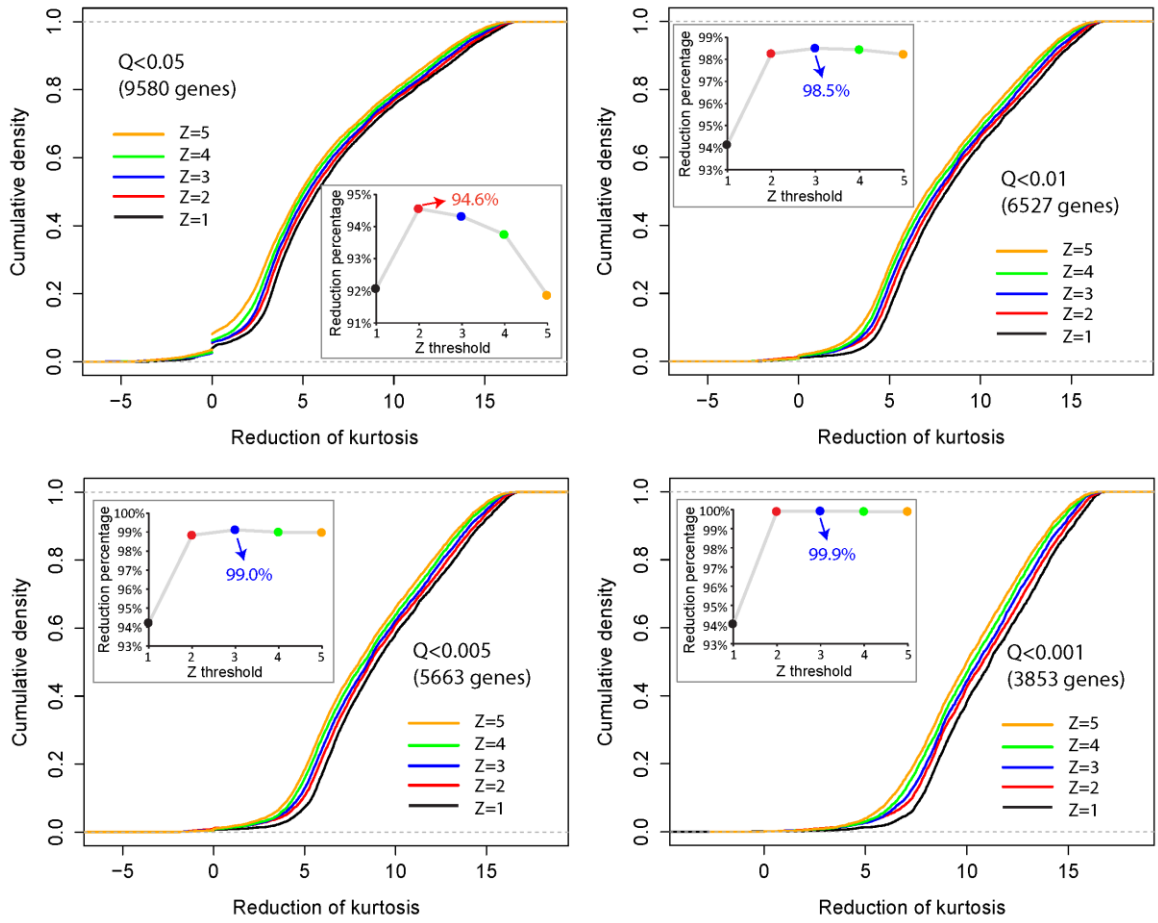

Figure S13. Implementing the reduction of Kurtosis analysis to infer genes with high tissue-specificity. Cumulative distribution of genes cross-tissue expression generated for different q-value (Q) thresholds obtained from R qvalue package[3]. The analysis is based on the Anscombe test for kurtosis using the Anscombe.test function in the R moments package as described in Li et al.[4] and in the method section. The x axis reports on the Kurtosis reduction when a certain Z threshold (ZMAD-normalized expression values) was applied, the y axis reports on the corresponding cumulative density. The insert panels correspond to the Kurtosis reduction percentage of genes when a given Z threshold was selected. Different colors denote different Z threshold and the corresponding cumulative curves. Z=3 was selected as the threshold to extract genes with leptokurtic behavior.

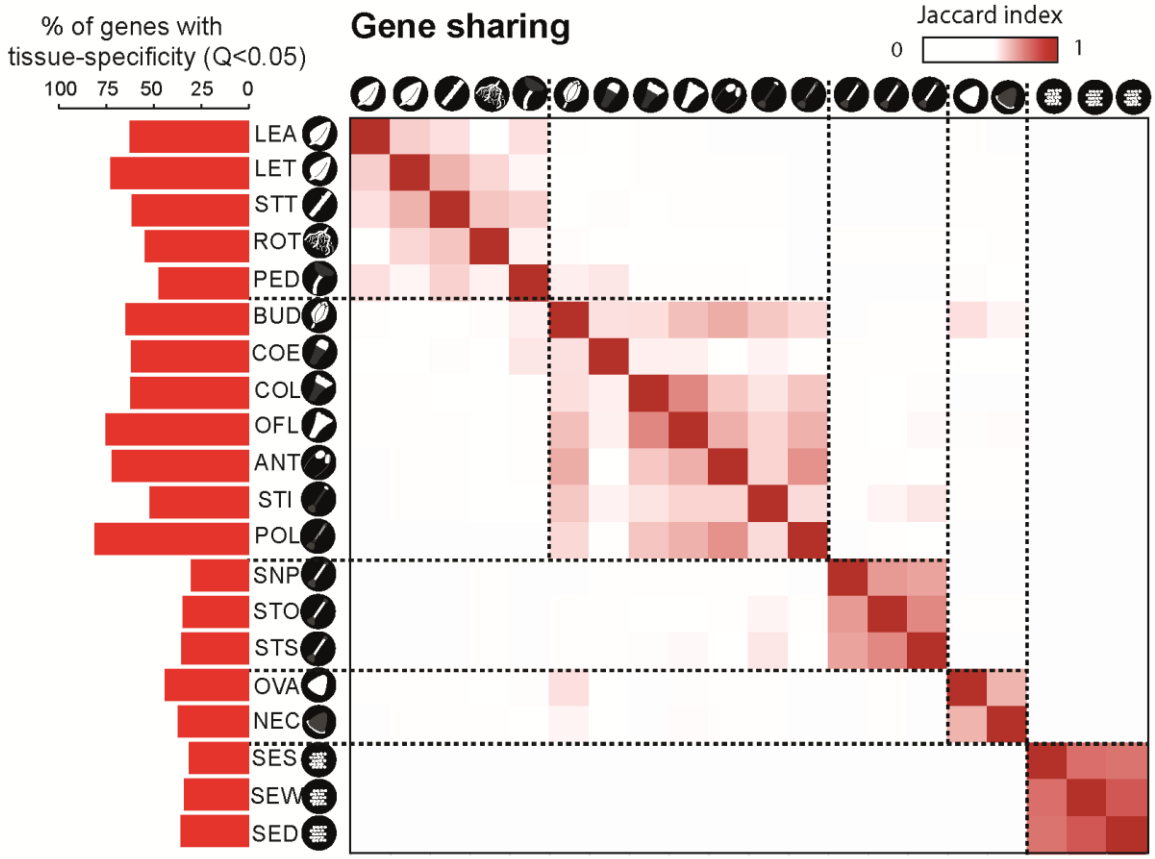

Figure S14. Gene-tissue specificity and gene-sharing among tissues of the transcriptome data-set. Left bar chart, % of genes showing tissue-specificity (Q<0.05) per tissue. Right Heatmap matrix visualizes genes sharing between tissues as measured using Jacquard index. LEA, leaves; LET, leaves treated with *Manduca sexta* oral secretion (OS); STT, stem from plants with leaves treated with OS; ROT, root from plants with leaves treated with OS; PED, pedicels; BUD, flower buds; NCO, non-matured corolla collected after protrusion from the calyx; COL, matured corolla collected after protrusion from the corolla; OFL, open flowers; ANT, anthers; STI, stigma; POL, pollen tubes; SNP, style without pollination; STO, style outcross-pollinated (2h after pollination); STS, style self-pollinated (2h after pollination); OVA, ovary; NEC, nectary; SES, seeds treated with liquid smoke; SEW, watered seeds; SED, matured seeds.

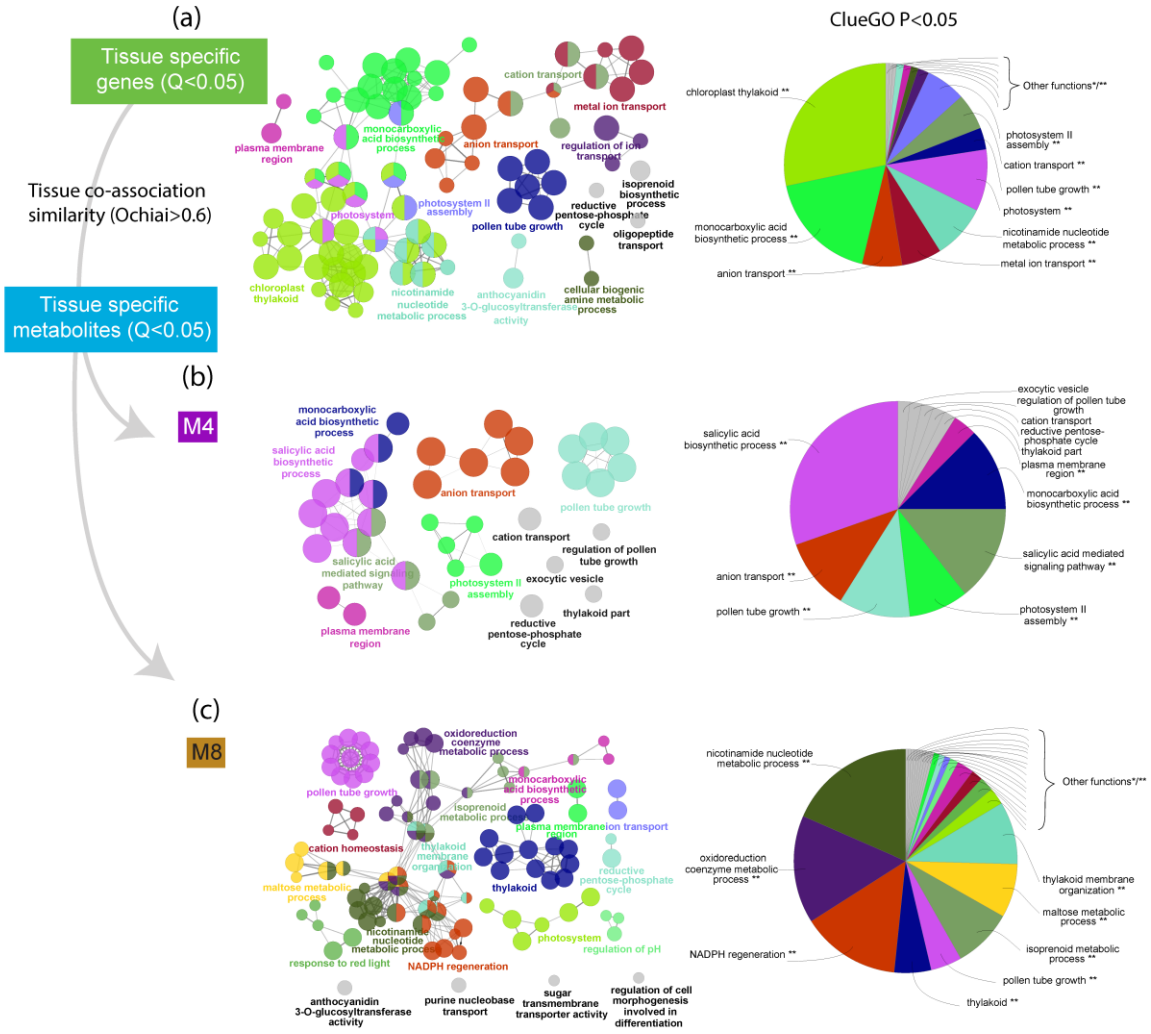

Figure S15. GO enrichment analysis for the pool of genes exhibiting tissue-specificity and co-tissue association with idMS/MSs classified as parts of modules M4 and M8. GO enrichment was generated via GlueGO[5]. The significant GO enrichment was depicted in overview pie charts as well as functionally grouped networks with terms as nodes and edges linked based on their kappa score level (>0.4), where only the label of the most significant term per group is shown. Node size is in proportion to the term enrichment significance and functionally related groups partially overlap. **(a)** Significant GO enrichment obtained for all tissue-specific genes (Q<0.05) from transcriptomic data. **(b)** and **(c)** GO enrichment from transcriptomic data sharing tissue co-associations (Ochiai>0.6) with the tissue-specific metabolites (Q<0.05) in module 4 and 8.

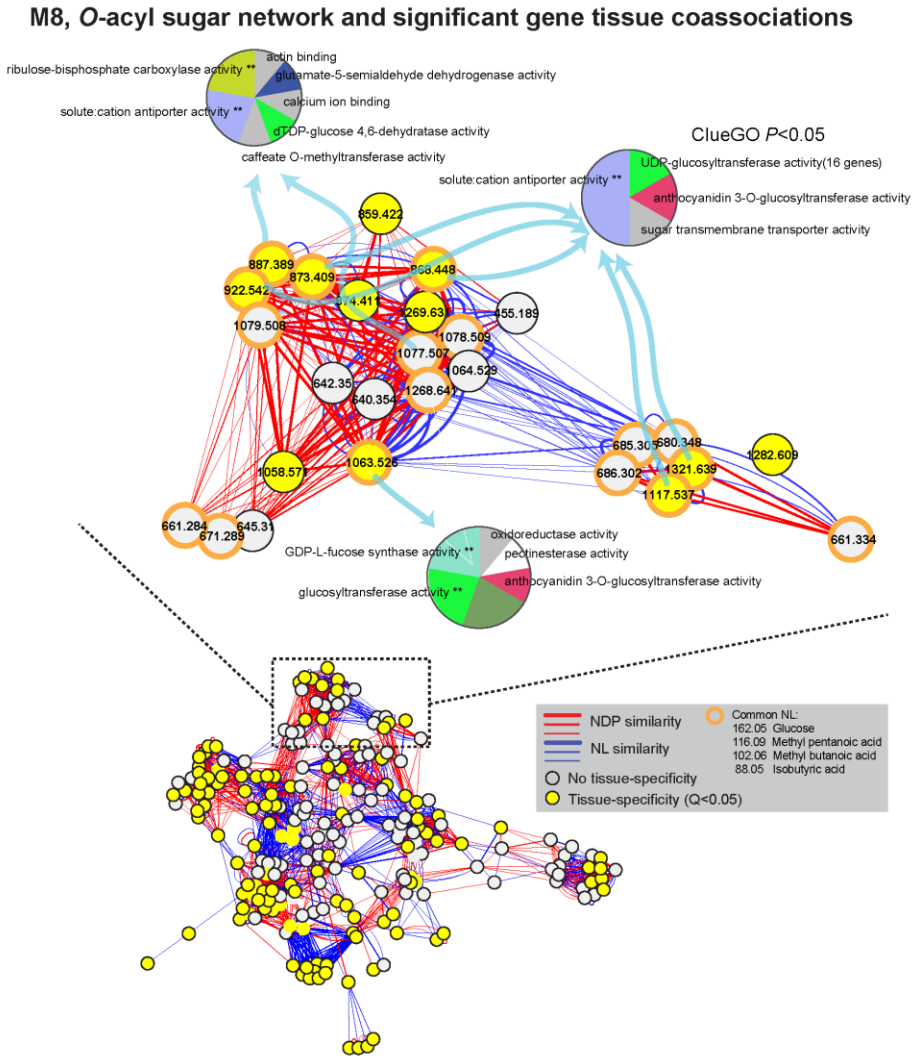

Figure S16. A sub-group of module 8 enriched in *O*-acyl sugar metabolites and GO terms from transcriptomic data sharing significant tissue co-associations with targeted metabolites. Molecular network constructed for module 8 which is enriched in *O*-acyl sugars is presented at bottom. Nodes represent idMS/MS spectra and edges correspond to structural similarities based on the two score types (NDP similarity calculated based on shared fragments between spectra and NL similarity calculated based on shared common neutral losses between spectra). Tissue-specificity is mapped onto the molecular network with node colors. Nodes that share typical *O*-acyl sugars neutral losses of glucose, methyl pentanoic acid, methyl butanoic acid and isobutyric acid are additionally circled in apricot. A zoom-in of the

network depicts metabolite-to-gene tissue co-association calculated as Ochiai similarity. GO terms were generated by GlueGO[5] from transcriptomic data sharing significant tissue co-associations with targeted metabolites.

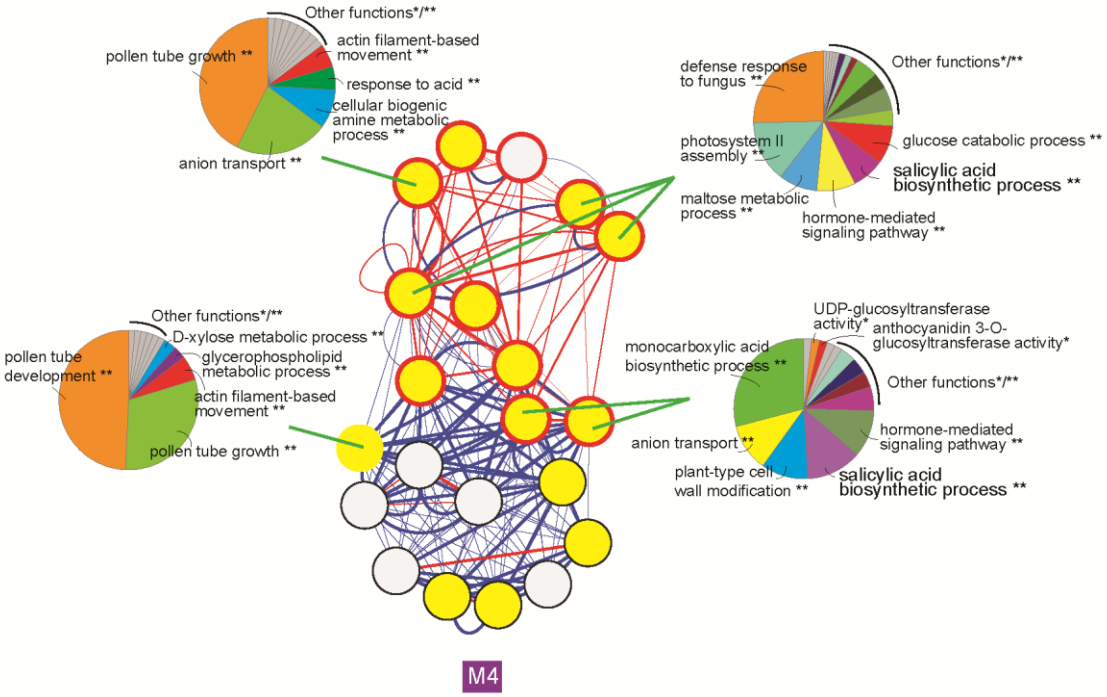

Figure S17. A sub-group of module 4 enriched in flavonoids and GO terms from transcriptomic data sharing significant tissue co-associations with targeted metabolites. Molecular network constructed for module 4 which originated from Figure 5 is presented. GO pie charts were generated by GlueGO[5] from transcriptomic data sharing significant tissue co-associations with given metabolites. Only significant GO terms are shown.

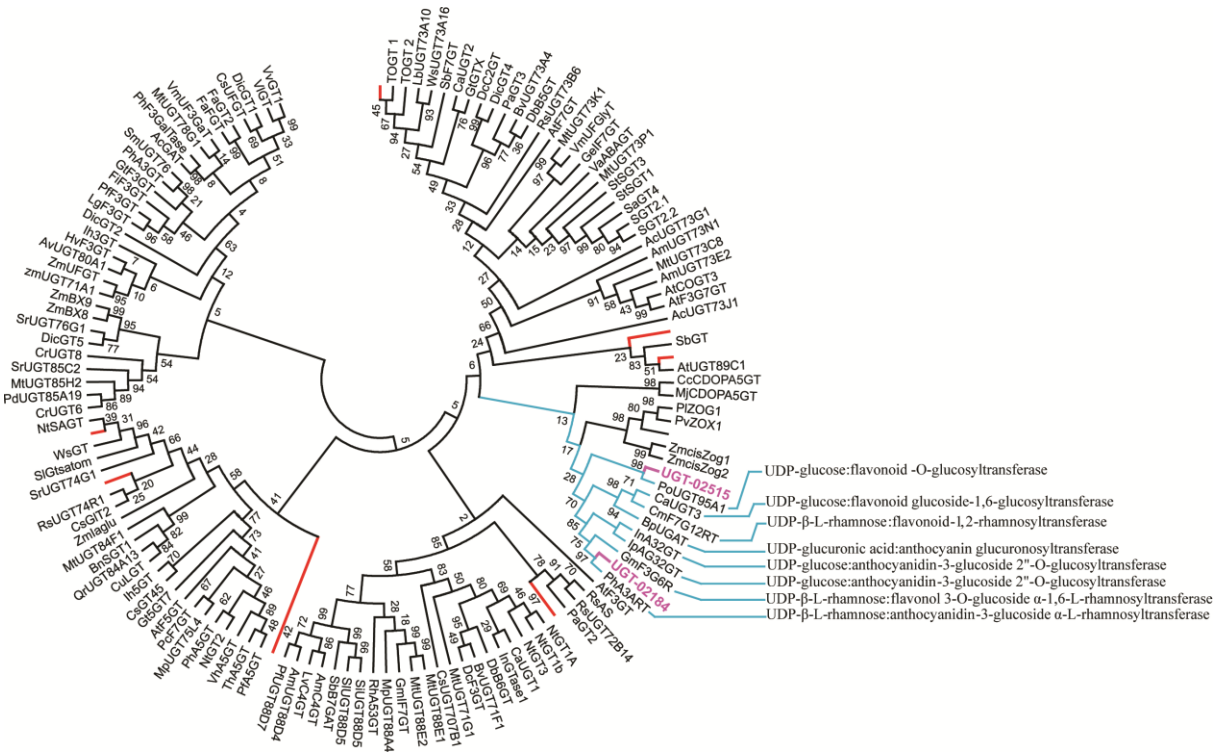

Figure S18. Phylogenetic tree analysis for the UDP-glucosyltransferases tested during the VIGS experiments. The tree was obtained by aligning characterized glycosyltransferases of the GT superfamily 1 (GT 1, 136 GT amino acid sequences) and inferring their phylogenetic relationship using the Maximum Likelihood method (bootstrap = 1000) based on the JTT matrix-based model[6]. Evolutionary analyses were conducted in MEGA5[7]. UGT-A (UGT-02184) and UGT-B (UGT-02515) analyzed in the present study are highlighted with purple branches. The red marked branches represent 10 other putative GTs (no names reported) initially used for virus-induced gene-silencing in *N. attenuata*.

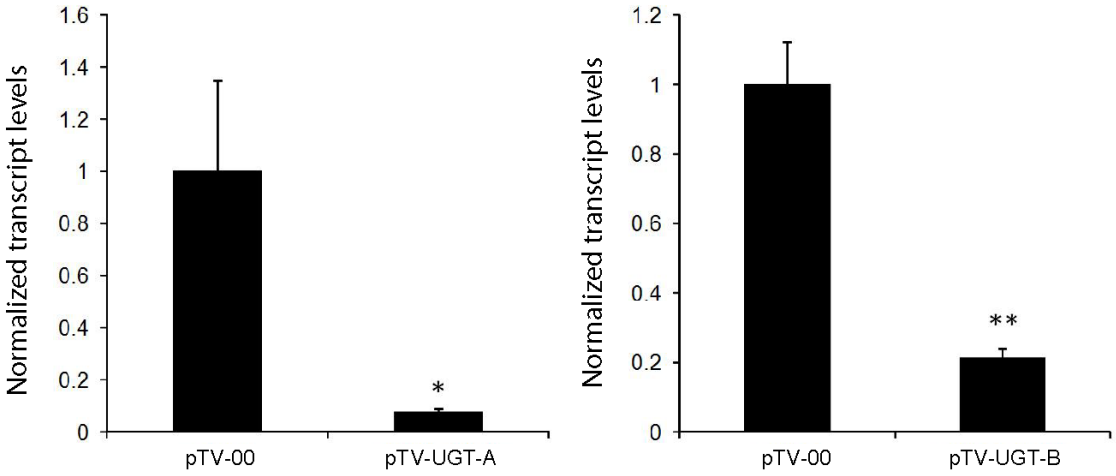

Figure S19. Gene silencing efficiency for the UDP-glucosyltransferases tested during VIGS experiments. Transcript abundances were determined in VIGS plants using gene specific primers and RT-qPCR after cDNA synthesis. Transcript levels (left panel, *pTV-UGT-A*; right panel, *pTV-UGT-B*) normalized to those of *ELONGATION FACTOR1* were determined in flower buds. Asterisks denote for significant differences between pTV-00 and UGT silenced lines (t-test, *$P < 0.05$, ** < 0.01). pTV-00, empty vector; pTV-UGT-A, UDP-glycosyltransferase-A silencing VIGS construct; pTV-UGT-B, UDP-glycosyltransferase-B silencing VIGS construct.

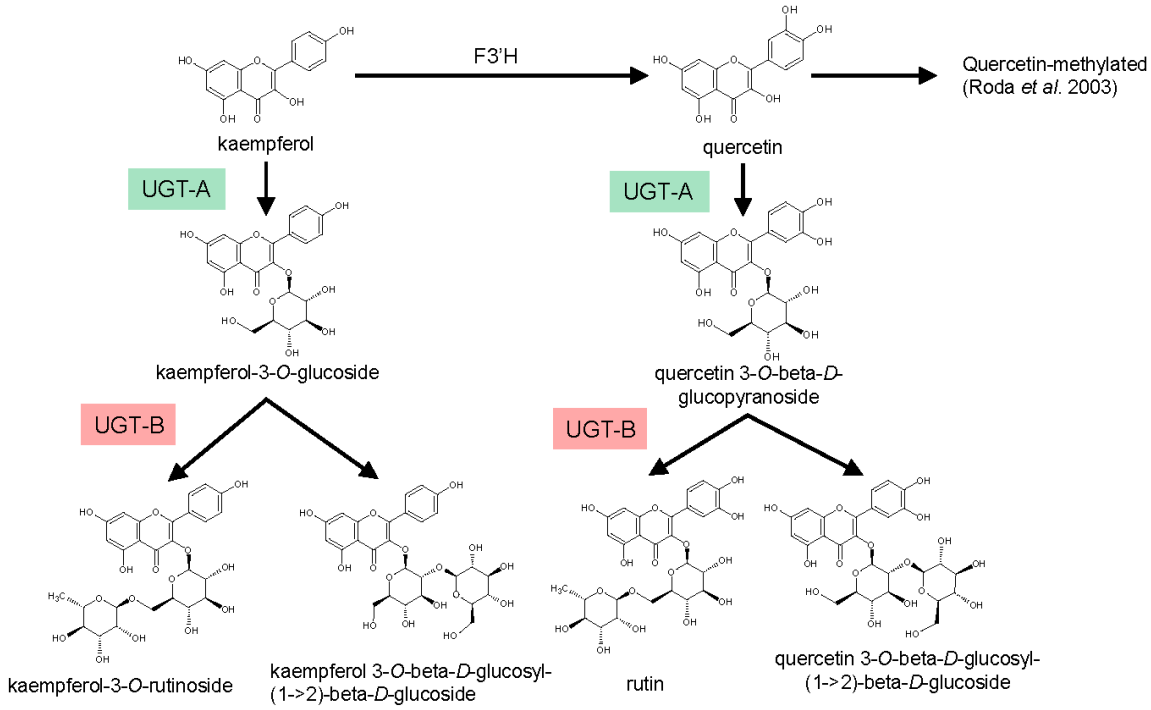

Figure S20. Proposed biosynthetic scheme for main flavonoid glycosides detected in this study. Scheme of the proposed flavonoid pathway with flavonoid structures previously reported in *Nicotiana attenuata* by Roda et al.[8]. UGT-A is proposed involving in glucosylation of flavonoid aglycones, UGT-B is proposed controlling the following rhamnosylation of the flavonoid glycosides. F3'H, flavonoid 3'-hydroxylase; UGT-A, UDP-glycosyltransferase-A; UGT-B, UDP-glycosyltransferase-B.

Table S1. List of the 52 neutral losses used for neutral loss similarity calculation

No.	Loss name	Loss formula	Mass difference ±
1	Alkane chains, waxes, fatty acids, methylation	CH_2	14.0157
2	Methane	CH_4	16.0313
3	Ammonium adduct/neutral ammonium loss	NH_3	17.0265
4	Water addition/loss	H_2O	18.0106
5	Salt adduct	K^+ to NH_4^+	20.9293
6	Salt adduct	Na^+ to H^+	21.9819
7	Ethine	C_2H_2	26.0157
8	Carbon monoxide	CO	27.9949
9	Ethene	C_2H_4	28.0313
10	Methylimine	CH_3N	29.0266
11	Formaldehyde	CH_2O	30.0106
12	Methylamine	CH_5N	31.0422
13	Sulfur	S	31.9721
14	Hydrogen sulfide	H_2S	33.9877
15	Salt adduct	K^+ to H^+	37.9559
16	Ketene	C_2H_2O	42.0106
17	Propylation	C_3H_6	42.0470
18	Cyanic acid	CHNO	43.0058
19	Carbon dioxide	CO_2	43.9898
20	Formic acid	CH_2O_2	46.0055
21	Butylation	C_4H_8	56.0626
22	Trimethylamine	C_3H_9N	59.0735
23	Acetic acid	$C_2H_4O_2$	60.0211
24	Urea	CH_4N_2O	60.0324
25	Sulfur dioxide	SO_2	63.9619
26	Pentene	C_5H_8	68.0626
27	Propionic acid	$C_3H_6O_2$	74.0368
28	Benzene	C_6H_6	78.0470
29	Sulfur trioxide	SO_3	79.9568
30	Malonic anhydride	$C_3H_2O_3$	86.0004
31	Isobutyric acid	$C_4H_8O_2$	88.0517
32	Putrescine	$C_4H_{12}N_2$	88.1000
33	Sulfuric acid	H_2SO_4	97.9674
34	Phosphoric acid	H_3PO_4	97.9769
35	Methyl butanoic acid	$C_5H_{10}O_2$	102.0618
36	Malonic acid	$C_3H_4O_4$	104.0110
37	Methyl pentanoic acid	$C_6H_{12}O_2$	116.0861
38	Dihydrogen vinyl phosphate	$C_2H_5O_4P$	123.9926
39	Pentose equivalent	$C_5H_8O_4$	132.0423

40	Spermidine	$C_7H_{19}N_3$	145.1579
41	Deoxy-Hexose equivalent	$C_6H_{10}O_4$	146.0579
42	Hexose-H2O	$C_6H_{10}O_5$	162.0528
43	Deoxy-Hexose equivalent	$C_6H_{12}O_5$	164.0685
44	Glucuronic acid-H2O	$C_6H_8O_6$	176.0321
45	Hexose	$C_6H_{12}O_6$	180.0634
46	Glucuronic acid	$C_6H_{10}O_7$	194.0427
47	Acyl-fructose	$C_8H_{12}O_6$	204.0655
48	Sinapic acid - H2O	$C_{11}H_{10}O_4$	206.0579
49	Glutathione+O-H2O	$C_{10}H_{15}N_3O_6S$	305.0682
50	Glutathione	$C_{10}H_{17}N_3O_6S$	307.0838
51	Sucrose-H2O	$C_{12}H_{20}O_{10}$	324.1057
52	Sucrose	$C_{12}H_{22}O_{11}$	342.1162

Table S2. List of primers used for qRT-PCR and gene fragment cloning for VIGS

Name	Sequence: (5' to 3') (for VIGS)	Sequence: (5' to 3') (for silencing efficiency)	Sequence: (5' to 3') (housekeeping elongation factor)
UGT-A forward	GCGGCGCTCGAGGTTGAGCATTATACTAAGGTGC	GACGCTAGAAGGAGTTTCAGG	TGGTATGGTTAAGATGCTTCCC
UGT-A reverse	GCGGCGGGATCCCAGGCAACCAATCTTCGTTGTC	CCACTGAATCGAACCAACAC	TGTCAACGCTCTTGATAACAC
UGT-B forward	GCGGCGCTCGAGGTGGTCCTACTGTATATGACCG	GGGAATTATTCATTCAGGTTGGG	TGGTATGGTTAAGATGCTTCCC
UGT-B reverse	GCGGCGGGATCCGGTAGCCCAGTTTGCTCCAGAC	GGCAACATAACCACTTGACAG	TGTCAACGCTCTTGATAACAC

Supplemental Methods

To reduce false positive errors resulting from spurious correlations from background noise due to the fact that some *m/z* features are only detected in a few samples, we compared data processing results obtained with and without the “fill peaks” function of XCMS (use for background noise correction) and calculated a background noise value from the average correction estimate used by this function to replace “NA” not detected peak intensities. When the “fill peaks” function is used, there still were many “0” intensity values in the dataset which affect the calculation of correlations, and these were replaced with the calculated background value. We also only considered features with intensities that were more than 3 times the background value and considered these as “true peaks”. Only *m/z* signals with at least six “true peaks” for the 28 samples precursors (MS1) and fragments datasets were considered for PCC calculation.

A precursor mass feature is further defined if its intensity across sample significantly correlate with the decreased intensity of the same mass feature subjected to low or high collision energies and that this feature is not annotated as an isotope peak by CAMERA. The correlation analysis was then conducted by calculating all possible precursor-to-product pairs within 9s – estimated maximum retention time window for peak deviation. Logically, *m/z* values for fragments should be lower than that of the precursor and MS/MS fragmentation should occur in the same sample position within the 28 sample dataset as the precursor from which it is derived.

Based on these two simple rules, we excluded assigned fragments at *m/z* values larger than that of the identified precursor as well as based the sample position for occurrence precursor and assigned fragments. Many in-source-fragmentation-generated mass features produced in the MS1 mode can also be selected as candidate precursors resulting in redundant compound idMS/MS. To reduce such data redundancy, we merged spectra if their NDP similarity exceeded 0.9 and they belong to the chromatographic “pcgroup” annotated by CAMERA. Finally we merged all the 4 collision energy results for precursor-to-fragment

associations into a final deconvoluted spectrum by choosing the highest intensity peak among all candidate peaks of the same *m/z* value at the different collision energies. This latter processing step is based on the composite spectrum concept and accounts for the different collision energy conditions required to maximize fragmentation possibilities since certain fragments are detected only at specific collision energies. After applying the entire pipeline and set of rules, 895 deconvoluted non-redundant spectra were reconstructed from the tissue-wide analysis.

REFERENCES FOR SUPPLEMENTAL MATERIAL

1 Gaquerel, E., Heiling, S., Schoettner, M., Zurek, G. & Baldwin, I. T. Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass spectrometry method for induced changes in *Nicotiana attenuata* leaves during simulated herbivory. *Journal of agricultural and food chemistry* **58**, 9418-9427 (2010).

2 Heiling, S., Khanal, S., Barsch, A., Zurek, G., Baldwin, I. T., Gaquerel, E. Using the knowns to discover the unknowns: MS-based dereplication uncovers structural diversity in 17-hydroxygeranyllinalool diterpene glycoside defense production in the Solanaceae. *The Plant Journal* **85**, 561-577 (2016).

3 Storey, J. D. A direct approach to false discovery rates. *J Roy Stat Soc B* **64**, 479-498 (2002).

4 Li, S. *et al.* Gene-sharing networks reveal organizing principles of transcriptomes in *Arabidopsis* and other multicellular organisms. *The Plant cell* **24**, 1362-1378 (2012).

5 Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-1093 (2009).

6 Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences : CABIOS* **8**, 275-282 (1992).

7 Tamura, K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-2739 (2011).

8 Roda, A. L., Oldham, N. J., Svatos, A. & Baldwin, I. T. Allometric analysis of the induced flavonols on the leaf surface of wild tobacco (*Nicotiana attenuata*). *Phytochemistry* **62**, 527-536 (2003).

4. Discussion

4.1. Translating some of the concepts of genomics into metabolomics

The successful development of sequencing techniques and sequence alignment tools in genomics has revolutionized biology. In the late 1980s and early 1990s, a wide array of algorithms for comparing nucleic acid sequences to infer genetic relatedness has emerged, the basic local alignment search tool (BLAST) standing for the most well-known and commonly used algorithm for the past 26 years (Altschul et al., 1990). The prosperity of homology-based alignment programs has been paralleled by exponentially explosive growth of sequencing techniques that are capable of sequencing millions of nucleic acids. The combination of the two has enabled biologist to query any organism's genomic architecture and to possibly annotate part of its genome. In the field of metabolomics, although we have witnessed a great advance in mass spectrometry with the range and accuracy of which is capable of generating millions of high resolution spectra from hundreds to thousands of samples, the depth of the "sequencing" in metabolomics comparing with that in genomics where the whole genome can be sequenced is still heavily constrained. Besides technological limitations, the constrains primarily arise due to the structural diversity and complexity of chemicals. As a consequence, usually multiple experimental extraction procedures were carried out to maximize the metabolic coverage. Additionally, data acquisition constrains limit the number of metabolites recruited for tandem MS/MS fragmentation. Meanwhile, computational approaches analogous to BLAST are far lagging. Types of tools and workflows that can align similar spectra and annotate metabolites of interest on a large-scale have just sporadically arisen in recent years (Watrous et al., 2012; Dührkop et al., 2015; Li et al., 2015; Allard et al., 2016). The success of genomics has pointed out the directions for metabolomics, but metabolomics are still nearly three decades left behind in the creation of straightforward tools and workflows which can harness massive spectra information and

finally revolutionize the way we look at metabolomics.

The objective of this thesis was to systematically explore secondary metabolism diversity in *N. attenuata*. The workflow implemented in this study was designed to fulfill three main purposes. The first one relates to the large-scale acquisition of MS/MS spectra information in a data-independent manner. The second purpose of this workflow is the straightforward processing of MS signatures to formulate more rapidly structural hypotheses on secondary metabolites of interest. This is a critical step as metabolite annotation, and in the best case its elucidation of secondary metabolites is a prerequisite for the functional characterization of metabolic systems and their related genes. In **manuscript I and III**, I established an integrative pipeline using an indiscriminant MS/MS approach for the holistic acquisition and fast exploration of population-level and within-plant-level known metabolites as well as unknown ones. The third purpose of the workflow presented is to functionally characterize the regulatory mechanism underlying the secondary metabolism. In **manuscript II**, I review the polymorphism in the jasmonate signaling and how this variation in the signaling regulatory cascade can be beneficial in detecting additional layer of secondary metabolites with defensive values. The variance among natural populations was used to link secondary metabolites and the upstream signaling regulation. In **manuscript III**, I notably showed that such variance analysis can be extended to the tissue scale to query causal genes. The correlational linkage between the expression of specific metabolic genes and characterized metabolic groups was built based on tissue co-association integrative analysis of transcriptomics and metabolomics. Candidate genes were further falsified using virus induced gene silencing approach. The systematic workflow described in this thesis on the one hand proposed a solution in exploring large-scale diverse secondary metabolites, on the other hand, bridged the missing link between causal genes and secondary metabolites (**Figure 1**). Taken together, the workflow presented provides the avenues traversing from the application of non-targeted data-independent MS/MS on the structural exploration of population-level and within-plant secondary metabolism leading to the gene function discoveries underlying the secondary metabolism diversity, which are described in rest of the discussion section.

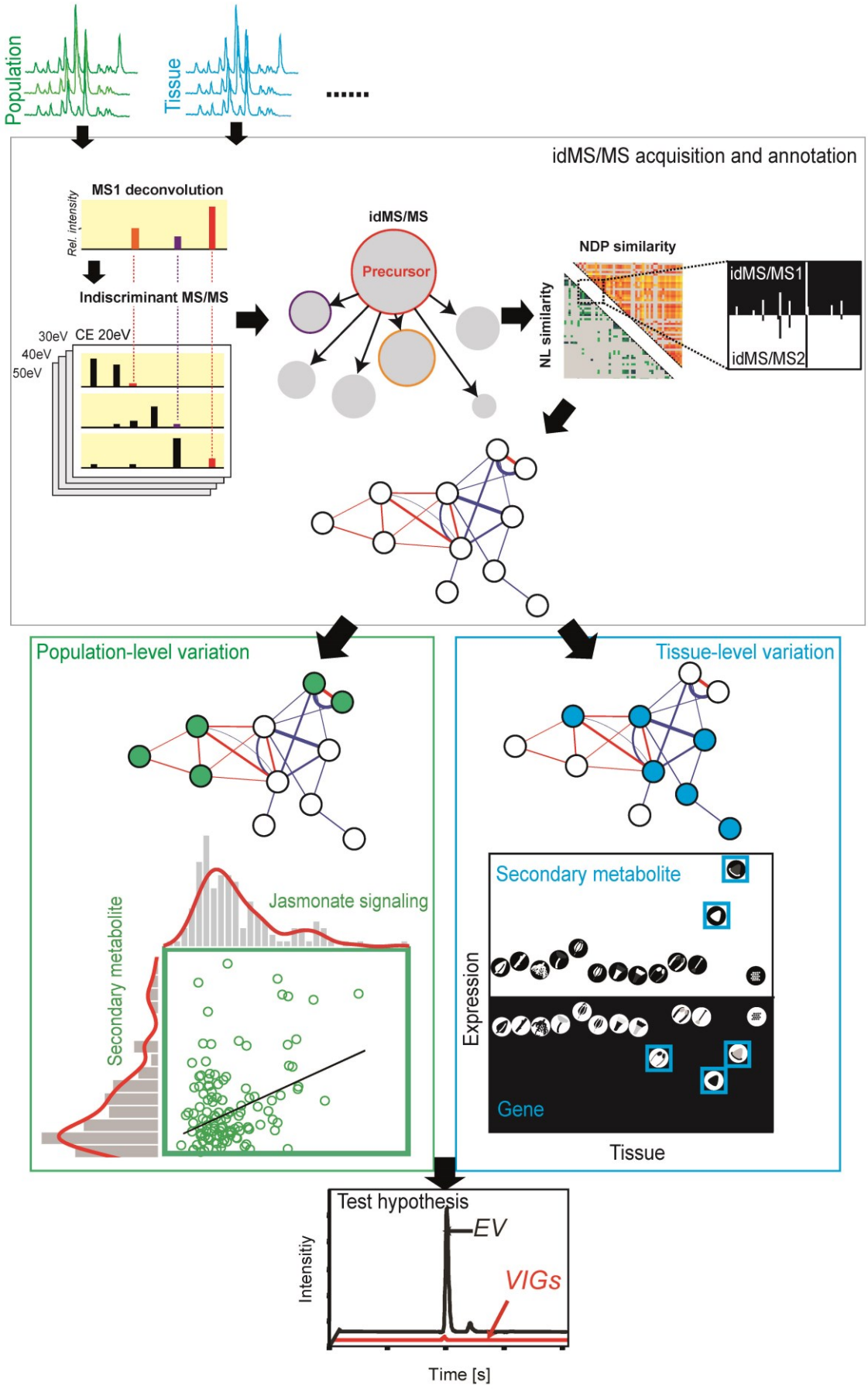
Population
Tissue
.....
idMS/MS acquisition and annotation
MS1 deconvolution
Rel. intensity
Indiscriminant MS/MS
30eV
CE 20eV
40eV
50eV
idMS/MS
Precursor
NDP similarity
NL similarity
idMS/MS1
idMS/MS2
Population-level variation
Secondary metabolite
Jasmonate signaling
Tissue-level variation
Secondary metabolite
Expression
Gene
Tissue
Test hypothesis
EV
VIGs
Intensity
Time [s]

Figure 1. The schematic workflow to systematic explore population-level and within-plant secondary metabolite diversity using MS/MS structural analysis. There are three main sections: (1) Creating idMS/MS dataset based on population-level and within-plant secondary metabolism variation. The resulting idMS/MS spectra were collected to calculate spectra similarity using both NDP and NL information for each idMS/MS pair. Using a bi-clustering approach, idMS/MSs were clustered into modules in which a molecular networking was carried out to annotate unknown metabolites. (2) Molecular networking was next used to facilitate the generation of hypotheses on regulatory mechanism underlying the metabolic diversity. (3) Finally, a reverse genetic approach was carried out to falsify the hypotheses.

4.2. Indiscriminant MS/MS acquisition and MS/MS structural analysis for rapidly formulating structural hypotheses on secondary metabolites

During the last decades, analytical approaches such as liquid chromatography-mass spectrometry (LC-MS) and NMR have tremendously advanced our understanding of the structure of small molecules. High resolution MS data when combined with molecular isotopic information is capable of generating molecular formulas of the analytes of high confidence (Kind and Fiehn, 2007). However, an MS approach alone cannot provide additional spectra information about molecular structures; isomers sharing the same molecular formula can not be resolved. Tandem MS data (MS/MS) offers many advantages in grasping structural insights of molecules, and thus is often applied as a prerequisite solution for structure identification. We are now in the transition phase from hypothesis-driven targeted metabolomics that are biased to certain secondary metabolite classes or well-mapped parts of central metabolism, to data-driven non-targeted metabolomics. The latter has shown to be extremely powerful in unraveling additional layers in metabolism in many studies (Ganna et al., 2014; Li et al., 2015; Ma et al., 2015).

The acquisition of non-targeted MS/MS data is categorized into data-dependent and data-independent methods. The data-independent acquisition offers great advantage in its coverage, however it is uninformative of precursor-fragment relationships, making it less

popular in application. Recently, a computational pipeline has been established to confidently assign precursor-fragment relationships (Broeckling et al., 2013). This offers the opportunity to reassess the data-independent acquisition method to obtain MS/MS spectra as large as possible. The workflow described in this thesis is based on indiscriminant/shotgun (data-independent) MS/MS (idMS/MS) data acquisition to collect a holistic repertoire of structural information on metabolic diversity. Shotgun MS/MS indiscriminately considers for fragmentation all signals within an *m/z* range set as large as possible (Broeckling et al., 2013; Broeckling et al., 2014). Although data-dependent MS/MS acquisition methods involving the selection of precursor ions for collision-induced dissociation (CID) fragmentation are more frequently applied, they suffered from several limitations in capturing the diversity of secondary metabolic. Firstly, the scan rate of the instrument limits the number of precursor selection for the further CID fragmentation in each acquisition cycle resulting in loss of information on unselected precursors, which decreases the comprehensiveness of the MS/MS analysis. Secondly, the *m/z* range for precursor isolation frequently brings in "contaminated" MS/MS data if the selection window is not set appropriately, leading to low-resolution performance. Thirdly, although the scan limitation can be circumvented by applying long repeated acquisition of the same sample and progressively shifting mass ranges selected for CID fragmentation to obtain as many spectra as possible (Matsuda et al., 2009), the procedure is extremely time consuming and it results in massive data redundancy resulting from iterative acquisition of MS/MS spectra of the same metabolite (Hoopmann et al., 2007; Broeckling et al., 2013). IdMS/MS approach, sometimes referred to as broad-scale MS/MS, on the other hand, has the main disadvantage of being uninformative about precursor-to-fragment relationships due to the unselecting procedure for precursors. However, once these relationships can be assigned confidently within the entirety of a metabolomics dataset, the idMS/MS method is extremely powerful in revealing new structural insights within the entire dataset (Broeckling et al., 2013; Broeckling et al., 2014). The precursor-to-fragment relationships were assigned using pearson correlation coefficient (PCC) analysis within the complete dataset. The low number of insufficient variable samples is

detrimental for the establishment of such relationships. Nevertheless, the confidential relationship assignment calculation would benefit from the variation of the secondary metabolites for the improvement of the resolution of PCC analysis and reciprocally, the holistic rich structure information captured by this approach will vastly facilitate the analysis of secondary metabolism diversity, which makes it ideal for such types of diversity analysis. In **manuscript I**, I used natural variation patterns for the correlation analysis. We have showed that idMS/MS approach can be efficiently implemented to older-generations' qTOF instruments by running replicated measurements of the same sample using idMS/MS at different collision energies (CEs) to maximize fragment coverage. New techniques using recent qTOF and triple TOF instruments will circumvent the coverage of fragmentation limitation with increased precision in ramped CIDs for a single run. In **manuscript III**, I extended this approach to the analysis of cross-tissue secondary metabolic variation within a single plant and implemented new rules to resolve idMS/MS redundancy. Due to the in-source fragmentation during ionization, multiple pseudo-MS3 idMS/MS spectra were sometimes collected for single metabolites resulting in redundant spectra information, such as the case of HGL-DTGs. I therefore calculated spectra similarity in the same retention time range and merged these pseudo-MS3 idMS/MS spectra into the so called composite idMS/MS spectra. A successful example was shown for the reconstruction of *N',N"*-caffeoyl,feruloylspermidine in which sub-idMS/MS of caffeoylspermidine and feruloylspermidine resulted from in-source fragmentation were sufficiently integrated into the composite idMS/MS spectra. Another valuable solution to increase the idMS/MS spectra integration performance would come from the integration of molecular formula information. With high resolution MS data combined with molecular isotopic information, it is now possible to calculate a relatively precise molecular formula, and this has already been applied to fragmentation tree studies (Rasche et al., 2011; Rasche et al., 2012). Such molecular formula information will aid the precursor-fragment assignment to better construct idMS/MS spectra.

Elucidating the structure of secondary metabolites is challenging. It is estimated that

presently, only 1.8% of spectra in an untargeted metabolomics experiment can be annotated (da Silva et al., 2015). The unknown secondary metabolites represent the absolute majority, the chemical signatures of which remain to be characterized. De novo elucidation of complete structures of these secondary metabolites is obviously mission impossible due to the tediousness of this task. It would require years of work with high-level expertise to isolate and identify the structure of even a single compound, not to mention that meanwhile mass spectrometers are already evolving rapidly for their fast acquisition and high resolution. The alternative solution which computationally annotates MS/MS spectra is inevitable. Tools to facilitate searching MS/MS databases to aid the annotation of unknown metabolites are a solution of rising potential. Particular progress has been made for certain metabolite classes, such as LipidBlast for lipids (Kind et al., 2013). More recently, a general strategy, CSI:FingerID which combines fragmentation tree computation and machine learning, was proposed to connect chemical structures found in public chemistry databases. While compared to shotgun proteomics which search peptide MS/MS data in peptide sequence database (Altelaar et al., 2013), MS/MS fragmentation in metabolomics exhibits large variation due to underlying chemical diversity, which makes it frequently hard to find a match from database searching. There are currently more than 60 million molecules stored in PubChem, yet 220,000 MS/MS spectra representing only about 20,000 molecules that are accessible for untargeted metabolomics (Johnson and Lange, 2015), the number of which accounts for less than one tenth of the estimated 200,000 metabolites. This is far from meeting the satisfaction point for non-targeted metabolomics searches. Recent efforts have been made to extend the database to *in-silico* MS/MS spectra in which MS/MS fragmentation was derived from theoretical prediction (Hufsky et al., 2014; Dührkop et al., 2015). As a result, a 10 fold increase in spectra was achieved compared to the total number of spectra present in all experimental MS/MS databases (Allard et al., 2016).

An alternative solution relies on intra-system MS/MS spectra matching. In-house databases are thus constructed to fulfil such a need (Matsuda et al., 2009; Matsuda et al., 2010). Within-dataset alignment of MS/MS spectra allows matching spectra in a

metabolomics experiment to molecules whose structure have not yet been elucidated and it simultaneously groups those structure-related compounds into compound classes. This is extremely powerful to maximize the usage of the limited existing structural knowledge in secondary metabolite exploration in non/new-model systems; specific compound classes are frequently missing in public databases for these secondary metabolites. This is especially the case for the ecological species *N. attenuata.* In **manuscript I**, I have shown that wild *N. attenuata* population induced a variety of secondary metabolites for its interactions with the specialist herbivore *Manduca sexta*. Query of these diverse idMS/MS spectra to Massbank was unsuccessful resulting in less than 18% hits of the total idMS/MS spectra with scores higher than 0.8 and frequently, highly scored idMS/MS spectra hits were frequently false positives not pointing to the right structures when manually analyzing the retrieved results. Although existing structural knowledge is limited, secondary metabolites within a species or closely-related species are frequently organized in a way that many compounds share the same backbone structure on which various modifications occur that diversify the secondary metabolites in a class. *Arabidopsis thaliana* possesses 40 glucosinolates (Kliebenstein et al., 2001), the diversity of which arise from elongation of the carbon side chain of the amino acid and subsequent transformations including hydroxylation, *O*-methylation, desaturation, glycosylation and acylation (Halkier and Gershenzon, 2006). 17-hydroxygeranyllinalool-diterpene glycosides (HGL-DTGs) in *N. attenuata* all share the same diterpene aglycone on which different glycosylation and malonylation take place at different positions (Heiling et al., 2010), and this knowledge has been successfully used to explore HGL-DTGs in closely-related *Nicotiana* species (Heiling, 2016). Additionally, the holistic acquisition of idMS/MS spectra in each biological sample provides sufficient metabolic space for the self-comparing needs so that spectra alignment can be easily made.

How to efficiently align MS/MS spectra is an ongoing challenge. MS/MS spectra alignment using cosine similarity or dot product has been long used in gas chromatography-mass spectrometry (GC-MS) to compare molecule fingerprints derived from electron ionization (EI) which results in compound-specific spectra with massive fragment

peaks to public databases. Such an approach has been transformed for comparing liquid chromatography–tandem mass spectrometry (LC-MS) data in many databases (Smith et al., 2005; Horai et al., 2010). A modified normalized dot product (NDP) was successfully applied to calculate similarities based on common fragments between MS/MS spectra, these studies focused on structurally complex microbial specialized metabolites that produced fragment-rich MS/MS spectra that are suitable for fragment-based MS/MS similarity alignment (Watrous et al., 2012). It is, however, less efficient when small analyzed molecules produce only a limited number of fragments. I have shown in **manuscript I** that many of the small molecules that I analyzed frequently produced only few fragments leading to poor spectra alignment performance. Common neutral losses (NLs) between MS/MS spectra which have been well-implemented in the context of MS/MS fragmentation tree studies can be used as another effective alignment factor for spectra similarity measurements (Rasche et al., 2011; Rasche et al., 2012). Two dimensional variables (NDP and NL) for spectra alignment are powerful but not easy to integrate. Supervised machine learning approaches such as Support Vector Machine (SVM) which was applied by CSI:FingerID, can efficiently handle such multi-variable data (Dührkop et al., 2015), however they require training datasets. These datasets are crucial for the development of search tools, since they aid prediction performance, and machine learning methods perform better with more comprehensive training sets. Yet this is not suitable for the non-model-species studies in which no such training set can be obtained. Unsupervised machine learning approaches which describe hidden structures from unlabeled data are qualified for the task. For example, hierarchical clustering, neural network and self-organizing map have been successfully applied to cluster metabolomics data (Hirai et al., 2005; Gulati et al., 2013; Hall et al., 2015). A recent bi-clustering approach named DiffCoEx which builds on the commonly used Weighted Gene Coexpression Network Analysis (WGCNA) framework for co-expression analysis was originally designed to find differentially co-expressed gene modules by considering two types of conditions (Tesson et al., 2010). Enlightened by this simple and sensitive method, I combined both NDP and NL variables into differential metabolic co-expression matrices, and

the resulting module clusters efficiently grouped either high NDP or NL-linked metabolites, or both. This procedure is essential in the pipeline as it clusters structure-related compounds into modules thus reducing the complexity in the metabolic space. With this reductionist approach, many unknown metabolites were grouped together with known ones in the same compound class, so that one can simply avoid dealing with complex MS/MS molecular networks produced from the overall dataset that are too complex to interpret. Most importantly, these compound modules frequently cluster metabolites sharing the same pathways, which facilitates the mining of pathway-specific compounds as shown for the annotation of uncharacterized compounds in the phenolic pathway in **manuscript I** and in the flavonoid pathway in **manuscript III**.

The next straightforward application of idMS/MS modules derived from bi-clustering was to fast formulate hypotheses about unknown metabolites based on the similarity of their idMS/MS spectra with that of known ones. Recently, various bioinformatics approaches have been developed to interpret large sets of MS/MS fragmentation data. One of the most innovative one was pioneered by the Dorrestein laboratory and involves the creation of MS/MS molecular networks (Watrous et al., 2012). This approach involves the generation of networks in which mass spectra (nodes) from one data-set or a subset of data are clustered using edges whose length is inversely proportional to the pairwise similarity between MS/MS spectra. Structurally more similar metabolites will be "pulled" into adjacent neighborhood than those dissimilar ones, this approach thus provides direct insights into unknown structures. I notably created molecular networking for certain module with edges linking idMS/MS spectra using both NDP and NL similarities. Analogues whose structure differ in modifications such as methylation or hydroxylation will reduce the performance of spectra similarity matching using only fragment-based approach due to the mass shifts. These analogues can then be efficiently "pulled" together by NL-based alignment. The visualization of the unknown regions of idMS/MS spectra in the networks enables structural hypotheses to be easily made. Most importantly, the network view of the metabolic groups also allows for the combined visualization of other biological information linked to a

given idMS/MS spectrum (in this thesis, I combined visualization of herbivory-induced natural variation and tissue-scale variation of secondary metabolites in the networks respectively), which can be virtually extended to any kind of biological information. A next interesting application would be mapping metabolite ontology information onto the network to better classify related compounds and define metabolic neighborhoods to increase annotation accuracy. Metabolic ontology information would greatly benefit metabolite annotation from the biological information about where a particular molecule or molecular family comes from and what it does. For instance, CSI:FinderID has weighted natural products and chemically synthesized compounds differentially in the scoring phase to avoid convoluting the search with synthetic compounds (Dührkop et al., 2015). There are already some databases with biologically relevant information available such as ChEBI (Hastings et al., 2013), which will be extremely beneficial for metabolic interpretation as the complexity of metabolic space expands.

4.3. Harnessing the power of statistical variance for gene discovery in secondary metabolism

The selection pressure imposed by insect herbivory has likely shaped many of the defense pathways that plants employ to maximize their fitness in their native habitats. In addition to constitutive defenses that repel herbivores, plants have evolved a more complex defense system that is specifically produced during insect feeding (Ballare, 2011). Many host plants of chewing insects have evolved the capacity to perceive elicitors in the insect's oral secretions (OS) which trigger signaling pathways and ultimately the synthesis of a repertoire of defensive proteins and secondary metabolites. The plasticity of secondary metabolite production in response to herbivory is largely mediated by jasmonic acid (JA) and its derivatives, the jasmonates (JAs). In **manuscript II**, I review within-plant and population-level heterogeneities in jasmonate signaling and dependent responses in the context of plant-insect interactions in *N. attenua.* I discuss means of manipulating jasmonate signaling for dissecting deviations from the canonical model of this hormonal pathway and

most importantly, how quantitative polymorphisms in jasmonate signaling can be analyzed for identifying additional layers of plant defense metabolites.

Natural variation has been extensively used to discover the function of genes. With current advances in sequencing and with the use of multi-parent advanced generation inter-cross (MAGIC) and recombinant inbred lines (RILs) combined with analytical approaches such as liquid chromatography-mass spectrometry (LC-MS) and NMR for targeted metabolic analysis, the exploration of natural variations in both model and crop species can now be used to identify the genetic bases of metabolic traits via quantitative traits locus mapping (Johal et al., 2008; Chan et al., 2011; Riedelsheimer et al., 2012). Large-scale metabolic quantitative trait loci (mQTL) analysis in combination with correlation-based network using natural variation has been applied to query genetic regions harboring particular secondary metabolic traits (Alseekh et al., 2015; Toubiana et al., 2015). These association-based approaches take advantages of existing variance in natural populations and are applied to query unknown secondary metabolites by correlating those with each other as well as with functionally characterized metabolic genes (Gong et al., 2013; Wen et al., 2014; Agerbirk et al., 2015; Kusano et al., 2015). These approaches can be used to infer associations between secondary metabolites and higher-order physiological information such as insect resistance (Kuzina et al., 2011). Similar approaches thus can be extended to the signaling level for the analysis of large-scale defense metabolism by correlating the variation across different genotypes with the aim of identifying naturally variable and jasmonate-dependent metabolic traits. To this end, I have notably observed that patterns of natural variation detected for herbivory-induced secondary metabolites only partly overlap with upstream variations in jasmonate production. This suggested that jasmonate-mediated metabolic accumulation mechanisms are under a complex regulatory system where additional components besides the canonical jasmonate signaling pathway might be involved. Most importantly, it illuminates many not-yet-characterized secondary metabolites with high natural diversity, which could provide predictive power for unraveling plant's jasmonate-mediated defensive arsenals.

Population geneticists have educated molecular biologists in how to harness the statistical power of variance arising from natural variation for gene pathway and function elucidations. The variance that occurs among tissues in a plant can be used for similar purposes. Previous tissue-level omics analyses have provided support to the claim that physiological differentiation is accompanied with qualitative variations of metabolic capacities (Uhlen et al., 2015). Recently, tissue-level co-expression analysis between genes and metabolites has been demonstrated to be an efficient approach for gene function analysis in secondary metabolism (Matsuda et al., 2010). In **manuscript III**, I have first used computational MS to accelerate the structural annotation of the diversity of compounds collected and then I notably addressed the question of regulatory mechanisms underlying secondary metabolism diversity using integrative analysis combining metabolomics with transcriptomics via tissue-co-association analysis. I have shown that this tissue-shared associations approach provides successful predictions for gene functions. The hypotheses regarding metabolic gene functions from metabolomics-transcriptomics integration were further falsified using an experimental gene-silencing approach. It is noteworthy that although the correlational analysis using variance represents a powerful approach for both metabolite and gene identification, the fundamental of the analysis - which relies on occurrence of single variance - is conditional. Metabolites that exhibit low or no variation will not be covered in such a type of analysis. Fortunately, many metabolites that fulfil key functions for environmental adaptations exhibit high natural variation due to the high selective pressures imposed on them. As such, this approach will be promising in unraveling fitness traits value for native populations.

4.4. Navigating metabolomics

Metabolomics, compared to genomics which has revolutionized biological research, is still in its infancy. Questions related to the nature of metabolic systems were frequently superficially addressed by biologists due to the complexity of metabolite identification. Currently, characterized secondary metabolism accounts for only a small fraction of the

metabolic space in plant kingdom. Of the 200,000 or so estimated plant compounds, only 25,000 are characterized as terpenoids, 12,000 are known as alkaloids and 8000 are reported as phenylpropanoids in literature (Croteau et al., 2000). The US National Cancer Institute has tested 35,000 plant extracts covering 12,000 to 13,000 plant species (collected by the United States Department of Agriculture) for anti-cancer activity. These surveys suggest that only 5% of plant species have had any kind of chemical analysis (Yonekura-Sakakibara and Saito, 2009). Numerous secondary metabolites with great biological potential are still not structurally identified and their ecological functions remain to be discovered. Yet due to the vast range of functional and structural diversity of secondary metabolites, the potential of mass spectrometry (MS)-based metabolomics and of the large-scale acquisition of tandem MS (MS/MS) spectra for many metabolites within a metabolic profile is severely constrained by the absence of straightforward classification and visualization pipelines so that metabolite interpretations can be easily made. Thus, I developed a metabolomics and computational pipeline using a non-targeted data-independent tandem mass spectrometry (MS/MS) approach combined with spectra similarity networking for the fast exploration of chemodiversity existing in secondary metabolism.

It is undisputed that MS/MS data alone are insufficient for full structural elucidation of unknown compounds, a task that is traditionally achieved after compound purification and de novo identification by NMR. Nevertheless, such an approach as described in this thesis will facilitate the process of formulating structural hypotheses by the combined interpretation of phenotypic information and mass spectrometric signatures. A recent breakthrough study has highlighted the power of metabolomics to predict biomarkers for the adjustments of the mammalian metabolome according to lifespan, phylogeny, and organ and lineage specialization (Ma et al., 2015). The workflow for the systematic exploration of metabolomics, and the concept of harnessing the statistical power of variance for the integration of genomics, transcriptomics and metabolomics for decoding gene function involved in secondary metabolism presented in this thesis should not be restricted to plants, since it can be extended to all other systems, such as microbials and animals, or even to the

metabolic navigation through different trophic levels where multi-systems interact with each other. Future directions will make use of genomics resources existing for related species of *N. attenuata* to extend the approach to the diagnosis of the gene divergence effects contributing most to species/tissue-level metabolic specialization.

Over the past decades, plant biologists have largely focused on quantitative and qualitative assessment of only limited collections of secondary metabolites, from only a few genotypes, raised in the rarefied environments of growth chambers, with little of the environmental complexity that sculpted the diversity of specialized metabolism. Questions on the mechanistic functions of genes underlying metabolic traits in the entire system have been largely overlooked or inaccessible to scientists. The future will require rapid and broad scale analysis of secondary metabolites and genes, within various environmental conditions and tissue types, to understand the entire regulatory system. Systematic metabolomics analysis is holding the key to assess such types of questions, to finally lead to a better understanding of the regulatory system in the field of biology.

References

Agerbirk N, Olsen CE, Heimes C, Christensen S, Bak S, Hauser TP (2015) Multiple hydroxyphenethyl glucosinolate isomers and their tandem mass spectrometric distinction in a geographically structured polymorphism in the crucifer Barbarea vulgaris. Phytochemistry **115:** 130-142

Allard PM, Peresse T, Bisson J, Gindro K, Marcourt L, Pham VC, Roussi F, Litaudon M, Wolfender JL (2016) Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. Anal Chem **88:** 3317-3323

Alseekh S, Tohge T, Wendenberg R, Scossa F, Omranian N, Li J, Kleessen S, Giavalisco P, Pleban T, Mueller-Roeber B, Zamir D, Nikoloski Z, Fernie AR (2015) Identification and mode of inheritance of quantitative trait loci for decondary metabolite sbundance in tomato. Plant Cell **27:** 485-512

Altelaar AF, Munoz J, Heck AJ (2013) Next-generation proteomics: towards an integrative

view of proteome dynamics. Nat Rev Genet **14:** 35-48

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:** 403-410

Ballare CL (2011) Jasmonate-induced defenses: a tale of intelligence, collaborators and rascals. Trends Plant Sci **16:** 249-257

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE (2014) RAMClust: A novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Analytical Chemistry **86:** 6812-6817

Broeckling CD, Heuberger AL, Prince JA, Ingelsson E, Prenni JE (2013) Assigning precursor-product ion relationships in indiscriminant MS/MS data from non-targeted metabolite profiling studies. Metabolomics **9:** 33-43

Chan EKF, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. PLoS Biol **9**

Croteau R, Kutchan TM, Lewis NG (2000) Natural Products (Secondary Metabolites). *In* Biochemistry & Molecular Biology of Plants, B. Buchanan, W. Gruissem, R. Jones, Eds., American Society of Plant Physiologists, Rockville, MD, pp. 1250–1268

da Silva RR, Dorrestein PC, Quinn RA (2015) Illuminating the dark matter in metabolomics. Proc Natl Acad Sci U S A **112:** 12549-12550

Dührkop K, Shen H, Meusel M, Rousu J, Bocker S (2015) Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc Natl Acad Sci U S A **112:** 12580-12585

Ganna A, Salihovic S, Sundstrom J, Broeckling CD, Hedman AK, Magnusson PK, Pedersen NL, Larsson A, Siegbahn A, Zilmer M, Prenni J, Arnlov J, Lind L, Fall T, Ingelsson E (2014) Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. PLoS Genet **10:** e1004801

Gong L, Chen W, Gao Y, Liu X, Zhang H, Xu C, Yu S, Zhang Q, Luo J (2013) Genetic analysis of the metabolome exemplified using a rice population. Proc Natl Acad Sci U

S A **110:** 20320-20325

Gulati J, Kim SG, Baldwin IT, Gaquerel E (2013) Deciphering herbivory-induced gene-to-metabolite dynamics in *Nicotiana attenuata* tissues using a multifactorial approach. Plant Physiol **162:** 1042-1059

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol **57:** 303-333

Hall LM, Hill DW, Menikarachchi LC, Chen MH, Hall LH, Grant DF (2015) Optimizing artificial neural network models for metabolomics and systems biology: an example using HPLC retention index data. Bioanalysis **7:** 939-955

Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N, Muthukrishnan V, Owen G, Turner S, Williams M, Steinbeck C (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res **41:** D456-463

Heiling S, Khanal, S., Barsch, A., Zurek, G., Baldwin, I. T., Gaquerel, E. (2016) Using the knowns to discover the unknowns: MS-based dereplication uncovers structural diversity in 17-hydroxygeranyllinalool diterpene glycoside defense production in the Solanaceae. The Plant Journal **85:** 561-577

Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, Schneider B, Jassbi AR, Baldwin IT (2010) Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. Plant Cell **22:** 273-292

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. J Biol Chem **280:** 25590-25595

Hoopmann MR, Finney GL, MacCoss MJ (2007) High-speed data reduction, feature

detection and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass Spectrometry. Analytical Chemistry **79:** 5620-5632

Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Tanaka K, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T (2010) MassBank: a public repository for sharing mass spectral data for life sciences. Journal of Mass Spectrometry **45:** 703-714

Hufsky F, Scheubert K, Bocker S (2014) New kids on the block: novel informatics methods for natural product discovery. Nat Prod Rep **31:** 807-817

Johal GS, Balint-Kurti P, Well CF (2008) Mining and harnessing natural variation: a little MAGIC. Crop Science **48:** 2066-2073

Johnson SR, Lange BM (2015) Open-access metabolomics databases for natural product research: present capabilities and future potential. Front Bioeng Biotechnol **3:** 22

Kind T, Fiehn O (2007) Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics **8:** 105

Kind T, Liu KH, Lee do Y, DeFelice B, Meissen JK, Fiehn O (2013) LipidBlast in silico tandem mass spectrometry database for lipid identification. Nat Methods **10:** 755-758

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. Plant Cell **13:** 681-693

Kusano M, Yang Z, Okazaki Y, Nakabayashi R, Fukushima A, Saito K (2015) Using metabolomic approaches to explore chemical diversity in rice. Mol Plant **8:** 58-67

Kuzina V, Nielsen JK, Augustin JM, Torp AM, Bak S, Andersen SB (2011) Barbarea vulgaris linkage map and quantitative trait loci for saponins, glucosinolates, hairiness and resistance to the herbivore Phyllotreta nemorum. Phytochemistry **72:** 188-198

Li D, Baldwin IT, Gaquerel E (2015) Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. Proc Natl Acad Sci U S A **112:** E4147-4155

Ma S, Yim SH, Lee SG, Kim EB, Lee SR, Chang KT, Buffenstein R, Lewis KN, Park TJ, Miller RA, Clish CB, Gladyshev VN (2015) Organization of the mammalian metabolome according to organ function, lineage specialization, and longevity. Cell Metab **22:** 332-343

Matsuda F, Hirai MY, Sasaki E, Akiyama K, Yonekura-Sakakibara K, Provart NJ, Sakurai T, Shimada Y, Saito K (2010) AtMetExpress development: a phytochemical atlas of Arabidopsis development. Plant Physiol **152:** 566-578

Matsuda F, Yonekura-Sakakibara K, Niida R, Kuromori T, Shinozaki K, Saito K (2009) MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. Plant Journal **57:** 555-577

Matsuda F, Yonekura-Sakakibara K, Niida R, Kuromori T, Shinozaki K, Saito K (2009) MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. Plant J **57:** 555-577

Rasche F, Scheubert K, Hufsky F, Zichner T, Kai M, Svatos A, Bocker S (2012) Identifying the unknowns by aligning fragmentation trees. Analytical Chemistry **84:** 3417-3426

Rasche F, Svatos A, Maddula RK, Bottcher C, Bocker S (2011) Computing fragmentation trees from tandem mass spectrometry data. Analytical Chemistry **83:** 1243-1251

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proceedings of the National Academy of Sciences of the United States of America **109:** 8872-8877

Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G (2005) METLIN: a metabolite mass spectral database. Ther Drug Monit **27:** 747-751

Tesson BM, Breitling R, Jansen RC (2010) DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC Bioinformatics **11:** 497

Toubiana D, Batushansky A, Tzfadia O, Scossa F, Khan A, Barak S, Zamir D, Fernie AR, Nikoloski Z, Fait A (2015) Combined correlation-based network and mQTL analyses efficiently identified loci for branched-chain amino acid, serine to threonine, and proline metabolism in tomato seeds. Plant Journal **81:** 121-133

Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Ponten F (2015) Proteomics. Tissue-based map of the human proteome. Science **347:** 1260419

Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, van der Voort M, Pogliano K, Gross H, Raaijmakers JM, Moore BS, Laskin J, Bandeira N, Dorrestein PC (2012) Mass spectral molecular networking of living microbial colonies. Proceedings of the National Academy of Sciences of the United States of America **109:** E1743-E1752

Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J, Yan J (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat Commun **5:** 3438

Yonekura-Sakakibara K, Saito K (2009) Functional genomics for plant natural product biosynthesis. Natural Product Reports **26:** 1466-1487

Summary

Plants produce a repository of structurally diverse secondary metabolites which play myriad ecological roles for plants' survival and reproductive fitness in nature. The study of secondary metabolism diversity will profoundly advance our understanding of plant's adaptive traits that have been subjected to natural selection during evolution. The analysis of intraspecific variations in secondary metabolism, however, has lagged behind and frequently been biased toward central metabolism. The potential of mass spectrometry (MS)-based metabolomics and of the large-scale acquisition of tandem MS (MS/MS) spectra for as many metabolites as possible within a metabolic profile is severely constrained by the absence of straightforward classification and visualization pipelines so that secondary metabolite and the underlying pathway interpretations can be easily made. From a mechanistic standpoint, secondary metabolism diversity attributes to the occurrence of multiplicity of genes in plant genomes. Yet the majority of metabolic gene functions remain however unknown either because the metabolites that they produce are unknown or no significant association has yet been identified between the expression of specific metabolic genes and characterized metabolic groups or phenotypes of interest.

In this thesis, I presented a straightforward workflow for the systematic exploration of population and tissue-level diversity in plant secondary metabolism using MS/MS structural analysis. The workflow combines the strengths of data-independent MS/MS acquisition to capture secondary metabolism diversity as large as possible in a data-set, of computational MS to accelerate the structural annotation of the diversity of compounds collected and of experimental gene-silencing to falsify hypotheses regarding metabolic gene functions from metabolomics-transcriptomics integration. Using natural variation, we notably employed correlation analysis between jasmonate signaling and large-scale defense metabolism. The resulting correlation maps have spotlighted non-previously identified metabolic layers in a plant's jasmonate-mediated defensive arsenal. In tissue-level exploration of secondary

metabolite diversity, we notably showed that the statistical power of variance arising from natural variation for pathway and functional elucidations can be similarly adapted to gene function characterization using variance that occurs among tissues. Tissue-metabolite association maps between metabolomics and transcriptomics were thus created for the prediction of tissue-specific gene functions underlying metabolic signatures. Finally, we employed experimental gene-silencing to falsify hypotheses regarding metabolic gene functions, as shown by the gene-silencing-based characterization of two UDP-glycosyltransferases for flavonoid metabolism.

The workflow presented here is obviously not directly applicable to de novo unknown compound identification, a task that is traditionally achieved after compound purification and de novo identification by NMR. Additionally, profiling the entire metabolomes of a given sample remains technically challenging due to the structural complexity of secondary metabolites. Nevertheless, the workflow allows biologists to amortize the vast amount of data produced by modern MS instrumentation and most importantly for their quest to understand gene function.

Zusammenfassung

Pflanzen produzieren eine Vielzahl von strukturell unterschiedlichen Sekundärmetaboliten, die in der Natur wichtig für ihr Überleben und ihre reproduktive Fitness sind. Die Erforschung der Diversität der Sekundärmetabolite wird uns helfen die Evolution der pflanzlichen Anpassungsmechanismen an ihre Umwelt besser zu verstehen. Jedoch wurde die Untersuchung interspezifischer Variationen des Sekundärmetabolismus bisher größtenteils vernachlässigt. Durch die Abwesenheit geeigneter Klassifizierungs- und Visualisierungsmethoden können die auf Massenspektrometrie (MS) beruhenden Metabolomanalyseverfahren, inklusive der auf Tandem-MS (MS/MS) basierten Techniken, ihr volles Potential nicht ausschöpfen. Mechanistisch betrachtet, hängt die Vielfalt des Sekundärstoffwechsels mit der genetischen Vielfalt von Pflanzen zusammen. Bisher sind jedoch ein Großteil der Zusammenhänge zwischen Genen und Metaboliten noch unklar. Zum einen, weil die Stoffe selber noch nicht bekannt sind oder sie nicht mit spezifischen Genen in Verbindung gebracht werden konnten.

In der vorliegenden Doktorarbeit präsentiere ich eine geeignete Methode zur systematischen Erforschung der populations- und gewebsspezifischen Vielfalt von pflanzlichen Sekundärmetaboliten durch die Benutzung einer MS/MS basierten Strukturanalyse. Hierbei werden die Stärken einer umfassenden, datenunabhängigen MS/MS Analyse, computergestützte Analyse der Massenspektren zur Strukturbestimmung und genetischen Verfahren zur Genstilllegung kombiniert um Hypothesen bezüglich des Zusammenspiels von Transkriptom und Metabolom zu überprüfen. Unter Ausnutzung der natürlichen Vielfallt haben wir Korrelationsanalysen zwischen dem Jasmonsäuresignalweg und einer Vielzahl von Verteidigungsmetaboliten durchgeführt. Die resultierenden Korrelations-Graphen geben einen Einblick in bisher unbekannte Sektoren der Jasmonsäure-gesteuerten Pflanzenverteidigung. Durch die gewebsspezifische Analyse des Sekundärstoffwechsels konnten wir zeigen, dass neben den natürlichen Variationen zwischen

Pflanzen auch die Unterschiede zwischen verschiedenen Pflanzengeweben genutzt werden können um ähnliche Genfunktionsanalysen durchzuführen. Mit Hilfe von Korrelationsgraphen aus Transkriptom- und Metabolomdaten verschiedener Pflanzengewebe erstellten wir Vorhersagen über gewebsspezifische Genfunktionen. Abschließend nutzen wir das gezielte Ausschalten einzelner Genen um potentielle Zusammenhänge mit bestimmten Metaboliten zu überprüfen. Dieses zeigen wir am Beispiel von zwei stillgelegten UDP-Glycosyltransferasen des Flavonoidstoffwechsels.

Die hier vorgestellte Methode ist nicht direkt auf die *de novo* Bestimmung von unbekannten chemischen Verbindungen anwendbar. Diese Aufgabe wird traditionell durch Aufreinigung der Verbindung und Analyse mittels NMR durchgeführt. Darüber hinaus bleibt es aufgrund der strukturellen Komplexität der Sekundärmetabolite immer noch eine anspruchsvolle Aufgabe ein möglichst umfassendes Metabolit-Profil einer Probe zu erstellen. Dennoch ermöglicht die in dieser Arbeit beschriebene Methode Biologen die riesigen Datenmengen, moderner MS Instrumente auszuwerten und damit die Funktion bestimmter Gene besser zu verstehen.

Bibliography

Agerbirk N, Olsen CE, Heimes C, Christensen S, Bak S, Hauser TP (2015) Multiple hydroxyphenethyl glucosinolate isomers and their tandem mass spectrometric distinction in a geographically structured polymorphism in the crucifer Barbarea vulgaris. Phytochemistry **115:** 130-142

Allard PM, Peresse T, Bisson J, Gindro K, Marcourt L, Pham VC, Roussi F, Litaudon M, Wolfender JL (2016) Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. Anal Chem **88:** 3317-3323

Alseekh S, Tohge T, Wendenberg R, Scossa F, Omranian N, Li J, Kleessen S, Giavalisco P, Pleban T, Mueller-Roeber B, Zamir D, Nikoloski Z, Fernie AR (2015) Identification and mode of inheritance of quantitative trait loci for decondary metabolite sbundance in tomato. Plant Cell **27:** 485-512

Altelaar AF, Munoz J, Heck AJ (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet **14:** 35-48

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:** 403-410

Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W, Baginsky S (2008) Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. Science **320:** 938-941

Bahulikar RA, Stanculescu D, Preston CA, Baldwin IT (2004) ISSR and AFLP analysis of the temporal and spatial population structure of the post-fire annual, *Nicotiana attenuata*, in SW Utah. BMC Ecol **4:** 12

Baldwin IT (1998) Jasmonate-induced responses are costly but benefit plants under attack in native populations. Proceedings of the National Academy of Sciences of the United States of America **95:** 8113-8118

Baldwin IT, Morse L (1994) Up in smoke: II. Germination of *Nicotiana attenuata* in response to smoke-derived cues and nutrients in burned and unburned soils. Journal of Chemical Ecology **20:** 2373-2391

Baldwin IT, Staszakkozinski L, Davidson R (1994) Up in smoke: I. Smoke-derived germination cues for postfire annual, *Nicotiana attenuata* torr. Ex. Watson. Journal of Chemical Ecology **20:** 2345-2371

Ballare CL (2011) Jasmonate-induced defenses: a tale of intelligence, collaborators and rascals. Trends Plant Sci **16:** 249-257

Barton KE, Koricheva J (2010) The ontogeny of plant defense and herbivory: characterizing general patterns using meta-analysis. American Naturalist **175:** 481-493

Binns SE, Arnason JT, Baum BR (2002) Phytochemical variation within populations of *Echinacea angustifolia* (Asteraceae). Biochemical Systematics and Ecology **30:** 837-854

Bowles D (2002) A multigene family of glycosyltransferases in a model plant, *Arabidopsis thaliana*. Biochemical Society Transactions **30:** 301-306

Bravo-Monzon AE, Rios-Vasquez E, Delgado-Lamas G, Espinosa-Garcia FJ (2014) Chemical diversity among populations of *Mikania micrantha*: geographic mosaic structure and herbivory. Oecologia **174:** 195-203

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE (2014) RAMClust: A novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Analytical Chemistry **86:** 6812-6817

Broeckling CD, Heuberger AL, Prince JA, Ingelsson E, Prenni JE (2013) Assigning precursor-product ion relationships in indiscriminant MS/MS data from non-targeted metabolite profiling studies. Metabolomics **9:** 33-43

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. Phytochemistry **62:** 471-481

Carreno-Quintero N, Acharjee A, Maliepaard C, Bachem CW, Mumm R, Bouwmeester

H, Visser RG, Keurentjes JJ (2012) Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality. Plant Physiol **158:** 1306-1318

Chan EKF, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. PLoS Biol **9**

Croteau R, Kutchan TM, Lewis NG (2000) Natural Products (Secondary Metabolites). *In* Biochemistry & Molecular Biology of Plants, B. Buchanan, W. Gruissem, R. Jones, Eds., American Society of Plant Physiologists, Rockville, MD, pp. 1250–1268

D'Auria JC, Chen F, Pichersky E (2002) Characterization of an acyltransferase capable of synthesizing benzylbenzoate and other volatile esters in flowers and damaged leaves of *Clarkia breweri*. Plant Physiol **130:** 466-476

D'Auria JC, Gershenzon J (2005) The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. Curr Opin Plant Biol **8:** 308-316

da Silva RR, Dorrestein PC, Quinn RA (2015) Illuminating the dark matter in metabolomics. Proc Natl Acad Sci U S A **112:** 12549-12550

Desbrosses GG, Kopka J, Udvardi MK (2005) Lotus japonicus metabolic profiling: development of gas chromatography-mass spectrometry resources for the study of plant-microbe interactions. Plant Physiol **137:** 1302-1318

Diezel C, Kessler D, Baldwin IT (2011) Pithy protection: *Nicotiana attenuata*'s jasmonic acid-mediated defenses are required to resist stem-boring *Weevil* larvae. Plant Physiol **155:** 1936-1946

Dührkop K, Shen H, Meusel M, Rousu J, Bocker S (2015) Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc Natl Acad Sci U S A **112:** 12580-12585

Ehrlich PR, Raven PH (1964) Butterflies and plants - a study in coevolution. Evolution **18:** 586-608

Fernie AR (2007) The future of metabolic phytochemistry: Larger numbers of metabolites,

higher resolution, greater understanding. Phytochemistry **68:** 2861-2880

Fernie AR, Aharoni A, Willmitzer L, Stitt M, Tohge T, Kopka J, Carroll AJ, Saito K, Fraser PD, DeLuca V (2011) Recommendations for reporting metabolite data. Plant Cell **23:** 2477-2482

Ferreyra MLF, Rius SP, Casati P (2012) Flavonoids: biosynthesis, biological functions, and biotechnological applications. Frontiers in Plant Science **3**

Fiehn O, Wohlgemuth G, Scholz M, Kind T, Lee do Y, Lu Y, Moon S, Nikolau B (2008) Quality control for plant metabolomics: reporting MSI-compliant studies. Plant J **53:** 691-704

Fraenkel GS (1959) The raison d'etre of secondary plant substances; these odd chemicals arose as a means of protecting plants from insects and now guide insects to food. Science **129:** 1466-1470

Fritz RS, Simms EL (1992) Plant resistance to herbivores and pathogens: ecology, evolution, and genetics. University of Chicago Press

Fu FF, Xue HW (2010) Coexpression analysis identifies Rice Starch Regulator1, a rice AP2/EREBP family transcription factor, as a novel rice starch biosynthesis regulator. Plant Physiol **154:** 927-938

Ganna A, Salihovic S, Sundstrom J, Broeckling CD, Hedman AK, Magnusson PK, Pedersen NL, Larsson A, Siegbahn A, Zilmer M, Prenni J, Arnlov J, Lind L, Fall T, Ingelsson E (2014) Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. PLoS Genet **10:** e1004801

Ganzera M, Guggenberger M, Stuppner H, Zidorn C (2008) Altitudinal variation of secondary metabolite profiles in flowering heads of *Matricaria chamomilla* cv. BONA. Planta Medica **74:** 453-457

Gaquerel E, Stitz M, Kallenbach M, Baldwin IT (2013) Jasmonate signaling in the field, part II: insect-guided characterization of genetic variations in jasmonate-dependent defenses of transgenic and natural *Nicotiana attenuata* populations. Methods Mol Biol **1011:** 97-109

Gong L, Chen W, Gao Y, Liu X, Zhang H, Xu C, Yu S, Zhang Q, Luo J (2013) Genetic analysis of the metabolome exemplified using a rice population. Proc Natl Acad Sci U S A **110:** 20320-20325

Gonzalez-Ballester D, Casero D, Cokus S, Pellegrini M, Merchant SS, Grossman AR (2010) RNA-seq analysis of sulfur-deprived Chlamydomonas cells reveals aspects of acclimation critical for cell survival. Plant Cell **22:** 2058-2084

Gulati J, Kim SG, Baldwin IT, Gaquerel E (2013) Deciphering herbivory-induced gene-to-metabolite dynamics in *Nicotiana attenuata* tissues using a multifactorial approach. Plant Physiol **162:** 1042-1059

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol **57:** 303-333

Hall LM, Hill DW, Menikarachchi LC, Chen MH, Hall LH, Grant DF (2015) Optimizing artificial neural network models for metabolomics and systems biology: an example using HPLC retention index data. Bioanalysis **7:** 939-955

Haslam E (1986) Secondary metabolism - fact and fiction. Natural Product Reports **3:** 217-249

Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N, Muthukrishnan V, Owen G, Turner S, Williams M, Steinbeck C (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res **41:** D456-463

Heiling S, Khanal, S., Barsch, A., Zurek, G., Baldwin, I. T., Gaquerel, E. (2016) Using the knowns to discover the unknowns: MS-based dereplication uncovers structural diversity in 17-hydroxygeranyllinalool diterpene glycoside defense production in the Solanaceae. The Plant Journal **85:** 561-577

Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, Schneider B, Jassbi AR, Baldwin IT (2010) Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. Plant Cell **22:**

273-292

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. J Biol Chem **280:** 25590-25595

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proceedings of the National Academy of Sciences of the United States of America **104:** 6478-6483

Holeski LM, Hillstrom ML, Whitham TG, Lindroth RL (2012) Relative importance of genetic, ontogenetic, induction, and seasonal variation in producing a multivariate defense phenotype in a foundation tree species. Oecologia **170:** 695-707

Hoopmann MR, Finney GL, MacCoss MJ (2007) High-speed data reduction, feature detection and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass Spectrometry. Analytical Chemistry **79:** 5620-5632

Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Tanaka K, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T (2010) MassBank: a public repository for sharing mass spectral data for life sciences. Journal of Mass Spectrometry **45:** 703-714

Hufsky F, Scheubert K, Bocker S (2014) New kids on the block: novel informatics methods for natural product discovery. Nat Prod Rep **31:** 807-817

Johal GS, Balint-Kurti P, Well CF (2008) Mining and harnessing natural variation: a little MAGIC. Crop Science **48:** 2066-2073

Johnson SR, Lange BM (2015) Open-access metabolomics databases for natural product

research: present capabilities and future potential. Front Bioeng Biotechnol **3:** 22

Jones CG, Firn RD (1991) On the evolution of plant secondary chemical diversity. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **333:** 273-280

Joyce AR, Palsson BO (2006) The model organism as a system: integrating 'omics' data sets. Nature Reviews Molecular Cell Biology **7:** 198-210

Kallenbach M, Bonaventure G, Gilardoni PA, Wissgott A, Baldwin IT (2012) Empoasca leafhoppers attack wild tobacco plants in a jasmonate-dependent manner and identify jasmonate mutants in natural populations. Proc Natl Acad Sci U S A **109:** E1548-1557

Karban R, Agrawal AA, Mangel M (1997) The benefits of induced defenses against herbivores. Ecology **78:** 1351-1355

Kaur H, Heinzel N, Schottner M, Baldwin IT, Galis I (2010) R2R3-NaMYB8 regulates the accumulation of phenylpropanoid-polyamine conjugates, which are essential for local and systemic defense against insect herbivores in *Nicotiana attenuata*. Plant Physiol **152:** 1731-1747

Kessler A, Halitschke R, Baldwin IT (2004) Silencing the jasmonate cascade: Induced plant defenses and insect populations. Science **305:** 665-668

Keurentjes JJ, Sulpice R (2009) The role of natural variation in dissecting genetic regulation of primary metabolism. Plant Signal Behav **4:** 244-246

Kim SG, Yon F, Gaquerel E, Gulati J, Baldwin IT (2011) Tissue specific diurnal rhythms of metabolites and their regulation during herbivore attack in a native tobacco, *Nicotiana attenuata*. Plos One **6:** e26214

Kind T, Fiehn O (2007) Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics **8:** 105

Kind T, Liu KH, Lee do Y, DeFelice B, Meissen JK, Fiehn O (2013) LipidBlast in silico tandem mass spectrometry database for lipid identification. Nat Methods **10:** 755-758

Kliebenstein DJ (2008) A role for gene duplication and natural variation of gene expression in the evolution of metabolism. Plos One **3:** e1838

Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T (2001) Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. Plant Physiol **126:** 811-825

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. Plant Cell **13:** 681-693

Kusano M, Yang Z, Okazaki Y, Nakabayashi R, Fukushima A, Saito K (2015) Using metabolomic approaches to explore chemical diversity in rice. Mol Plant **8:** 58-67

Kuzina V, Nielsen JK, Augustin JM, Torp AM, Bak S, Andersen SB (2011) *Barbarea vulgaris* linkage map and quantitative trait loci for saponins, glucosinolates, hairiness and resistance to the herbivore *Phyllotreta nemorum*. Phytochemistry **72:** 188-198

Li D, Baldwin IT, Gaquerel E (2015) Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. Proc Natl Acad Sci U S A **112:** E4147-4155

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics **154:** 459-473

Ma S, Yim SH, Lee SG, Kim EB, Lee SR, Chang KT, Buffenstein R, Lewis KN, Park TJ, Miller RA, Clish CB, Gladyshev VN (2015) Organization of the mammalian metabolome according to organ function, lineage specialization, and longevity. Cell Metab **22:** 332-343

Matsuba Y, Nguyen TT, Wiegert K, Falara V, Gonzales-Vigil E, Leong B, Schafer P, Kudrna D, Wing RA, Bolger AM, Usadel B, Tissier A, Fernie AR, Barry CS, Pichersky E (2013) Evolution of a complex locus for terpene biosynthesis in *solanum*. Plant Cell **25:** 2022-2036

Matsuda F, Hirai MY, Sasaki E, Akiyama K, Yonekura-Sakakibara K, Provart NJ, Sakurai T, Shimada Y, Saito K (2010) AtMetExpress development: a phytochemical atlas of Arabidopsis development. Plant Physiol **152:** 566-578

Matsuda F, Yonekura-Sakakibara K, Niida R, Kuromori T, Shinozaki K, Saito K (2009) MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. Plant Journal **57:** 555-577

Matsuda F, Yonekura-Sakakibara K, Niida R, Kuromori T, Shinozaki K, Saito K (2009) MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. Plant J **57:** 555-577

Milkowski C, Strack D (2004) Serine carboxypeptidase-like acyltransferases. Phytochemistry **65:** 517-524

Moco S, Capanoglu E, Tikunov Y, Bino RJ, Boyacioglu D, Hall RD, Vervoort J, De Vos RCH (2007) Tissue specialization at the metabolite level is perceived during the development of tomato fruit. Journal of Experimental Botany **58:** 4131-4146

Moore BD, Andrew RL, Kulheim C, Foley WJ (2014) Explaining intraspecific diversity in plant secondary metabolites in an ecological context. New Phytol **201:** 733-750

Mothes K (1955) Physiology of alkaloids. Annual Review of Plant Physiology and Plant Molecular Biology **6:** 393-432

Neumann S, Bocker S (2010) Computational mass spectrometry for metabolomics: Identification of metabolites and small molecules. Analytical and Bioanalytical Chemistry **398:** 2779-2788

Nomura T, Murase T, Ogita S, Kato Y (2015) Molecular identification of tuliposide B-converting enzyme: a lactone-forming carboxylesterase from the pollen of tulip. Plant Journal **83:** 252-262

Onkokesung N, Gaquerel E, Kotkar H, Kaur H, Baldwin IT, Galis I (2012) MYB8 controls inducible phenolamide levels by activating three novel hydroxycinnamoyl-coenzyme A:polyamine transferases in *Nicotiana attenuata*. Plant Physiol **158:** 389-407

Pelser PB, de Vos H, Theuring C, Beuerle T, Vrieling K, Hartmann T (2005) Frequent gain and loss of pyrrolizidine alkaloids in the evolution of *Senecio* section *Jacobaea* (Asteraceae). Phytochemistry **66:** 1285-1295

Quina FH, Moreira PF, Vautier-Giongo C, Rettori D, Rodrigues RF, Freitas AA, Silva PF, Macanita AL (2009) Photochemistry of anthocyanins and their biological role in plant tissues. Pure and Applied Chemistry **81:** 1687-1694

Rajniak J, Barco B, Clay NK, Sattely ES (2015) A new cyanogenic metabolite in *Arabidopsis* required for inducible pathogen defence. Nature **525:** 376-+

Ramakrishna A, Ravishankar GA (2011) Influence of abiotic stress signals on secondary metabolites in plants. Plant Signal Behav **6:** 1720-1731

Rasche F, Scheubert K, Hufsky F, Zichner T, Kai M, Svatos A, Bocker S (2012) Identifying the unknowns by aligning fragmentation trees. Analytical Chemistry **84:** 3417-3426

Rasche F, Svatos A, Maddula RK, Bottcher C, Bocker S (2011) Computing fragmentation trees from tandem mass spectrometry data. Analytical Chemistry **83:** 1243-1251

Rhoades DF, Cates RG (1976) Toward a general theory of plant antiherbivore chemistry. Recent Advances in Phytochemistry **10:** 168–213

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proceedings of the National Academy of Sciences of the United States of America **109:** 8872-8877

Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *PHASEOLUS VULGARIS*. Genetics **8:** 552-560

Schuler MA, Werck-Reichhart D (2003) Functional genomics of P450s. Annual Review of Plant Biology **54:** 629-667

Schuster J, Knill T, Reichelt M, Gershenzon J, Binder S (2006) BRANCHED-CHAIN AMINOTRANSFERASE4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. Plant Cell **18:** 2664-2679

Shigetomi K, Shoji K, Mitsuhashi S, Ubukata M (2010) The antibacterial properties of 6-tuliposide B. Synthesis of 6-tuliposide B analogues and structure-activity relationship. Phytochemistry **71:** 312-324

Shoji K, Ubukata M, Momonoi K, Tsuji T, Morimatsu T (2005) Anther-specific production of antimicrobial tuliposide B in tulips. Journal of the Japanese Society for Horticultural Science **74:** 469-475

Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G (2005) METLIN: a metabolite mass spectral database. Ther Drug Monit **27:** 747-751

Steppuhn A, Baldwin IT (2007) Resistance management in a native plant: nicotine prevents herbivores from compensating for plant protease inhibitors. Ecol Lett **10:** 499-511

Steppuhn A, Gase K, Krock B, Halitschke R, Baldwin IT (2004) Nicotine's defensive function in nature. PLoS Biol **2:** 1074-1080

Strauss SY, Agrawal AA (1999) The ecology and evolution of plant tolerance to herbivory. Trends Ecol Evol **14:** 179-185

Tesson BM, Breitling R, Jansen RC (2010) DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC Bioinformatics **11:** 497

Toubiana D, Batushansky A, Tzfadia O, Scossa F, Khan A, Barak S, Zamir D, Fernie AR, Nikoloski Z, Fait A (2015) Combined correlation-based network and mQTL analyses efficiently identified loci for branched-chain amino acid, serine to threonine, and proline metabolism in tomato seeds. Plant Journal **81:** 121-133

Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Ponten F (2015) Proteomics. Tissue-based map of the human proteome. Science **347:** 1260419

Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, van der Voort M, Pogliano K, Gross H, Raaijmakers JM, Moore BS, Laskin J, Bandeira N, Dorrestein PC (2012) Mass spectral molecular networking of living microbial

colonies. Proceedings of the National Academy of Sciences of the United States of America **109:** E1743-E1752

Weinhold A, Baldwin IT (2011) Trichome-derived *O*-acyl sugars are a first meal for caterpillars that tags them for predation. Proceedings of the National Academy of Sciences of the United States of America **108:** 7855-7859

Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J, Yan J (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat Commun **5:** 3438

Weng JK, Philippe RN, Noel JP (2012) The rise of chemodiversity in plants. Science **336:** 1667-1670

Wink M (1993) Allelochemical properties or the raison d'être of alkaloids. The Alkaloids **43:** 1-118

Wink M (2003) Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry **64:** 3-19

Yonekura-Sakakibara K, Saito K (2009) Functional genomics for plant natural product biosynthesis. Natural Product Reports **26:** 1466-1487

Züst T, Heichinger C, Grossniklaus U, Harrington R, Kliebenstein DJ, Turnbull LA (2012) Natural enemies drive geographic variation in plant defenses. Science **338:** 116-119

Curriculum Vitae

Dapeng Li

Born: 01.11.1986, Changchun, China
Nationality: China
Email: roclee1101@gmail.com
Gender: Male

Education

2012-2016: Ph. D. (Metabolomics)

Max-Planck-Institute for Chemical Ecology, Jena, Germany

Dissertation: "Tissue and population-level diversity in plant secondary metabolism: a systematic exploration using MS/MS structural analysis"

2009-2012: M. Sc (Bioinformatics)

Tongji University, Shanghai, China

Subject: "A novel structural position-specific scoring matrix for the prediction of protein secondary structures"

2005-2009: B. Sc (Analytical Chemistry)

Tongji University, Shanghai, China

Research Experience

2010-2011: Research assistant

Tongji University, Shanghai, China

Project: Propose for the first time a novel method for the prediction of protein secondary structure which outperformed state-of-the-art methods.

Funding: National Natural Science Foundation, Chinese National Program (973, 863 Program)

2009-2010: Research assistant

Tongji University, Shanghai, China

Project: Metabolomics analysis of prostate cancer cell of transgenic mice

Funding: National Natural Science Foundation

2008-2009: Research assistant

Tongji University, Shanghai, China

Project: Machine learning based on Conditional Random Fields

Funding: National Natural Science Foundation

Publication

- **Li D.**, Heiling S., Baldwin I. T., Gaquerel E. (2016). Illuminating a plant's tissue-specific metabolic diversity using computational metabolomics and information theory. **Proceedings of the National Academy of Sciences of the United States of America**, 113(47), E7610-E7618.
- Brockmöller T., Ling Z., **Li D.**, Gaquerel E. Xu S. (2016). Coyote tobacco data hub (CTDH): An integrative platform for Nicotiana attenuata genomic, transcriptomic and metabolomic data. (**In submission to BMC Genomics**)
- **Li D.**, Baldwin I. T., Gaquerel E. (2016). Beyond the canon: within-plant and population-level heterogeneity in jasmonate signaling engaged by plant-insect interactions. **Plants**. (In press)
- **Li D.**, Baldwin I. T., Gaquerel E. (2015). Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. **Proceedings of the National Academy of Sciences of the United States of America**, 112(30), E4147-E4155.
- Song Q., Li T., Cong P., Sun J., **Li D.**, Tang S. (2012). Predicting turns in proteins with a unified model. **PLoS ONE**, 7(11).
- **Li D.**, Li T., Cong P., Xiong W., Sun J. (2012). A novel structural position-specific scoring matrix for the prediction of protein secondary structures. **Bioinformatics**, 28(1).

- Zhu Y., Li T., **Li D.**, Zhang Y., Xiong W., et al. (2012). Using predicted shape string to enhance the accuracy of γ-turn prediction. **Amino acids**, 42(5).

Oral Presentations

- **Li D.** (2014). Navigating diversity in metabolism: a novel structural analysis of indiscriminant MS/MS profiles. Talk presented at Kick-off meeting for the Max-Planck-Partner-Group at the Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, CN (invited)

Poster Presentations

- Schäfer M., **Li D.**, Goldberg J.K., Baldwin I.T., Schuman M. (2015). Secondary metabolite flow through different trophic levels: Nicotine and related metabolites. Poster presented at SFB CHEMBIOSYS KICK-OFF Retreat, DFG, Saalfeld, DE
- **Li D.**, Heiling S., Gulati J., Schöttner M., Gaquerel E., Baldwin I.T. (2014). Connecting the transcriptome to the metabolome: from mass features to gene discovery. Poster presented at SAB Meeting 2014, MPI for Chemical Ecology, Jena, DE
- **Li D.**, Baldwin I.T., Gaquerel E. (2014). An indiscriminant MS/MS metabolomics approach unravels natural diversity patterns in defense responses of Nicotiana attenuata populations. Poster presented at 13th IMPRS Symposium, MPI for Chemical Ecology, Dornburg, DE
- **Li D.** (2013). An indiscriminant MS/MS approach unravels natural diversity patterns in defense responses of Nicotiana attenuata populations. Poster presented at 12th IMPRS Symposium, MPI for Chemical Ecology, Jena, DE

Language Skills

Chinese: Native speaker

English: Fluent

German: Basic

Acknowledgment

On the completion of my thesis, I would like to express my deepest gratitude to all those whose kindness and advice have made this work possible. I am greatly indebted to my advisors Professor **Ian T. Baldwin** who gave me valuable instructions and has improved me in language. I want to deeply thank you for the opportunity to work in this great institute with exciting projects, good resources and nice people, and of course, for your guidance, advice and support. Your effective advice, shrewd comments have kept the thesis in the right direction. I will never forget those scientific discussions with you when you can always found metaphors for better explanation for biological questions. Your way of thinking and enthusiasm in science have shaped me as a scientist. **Dr. Emmanuel Gaquerel** as my second supervisor, I would like to express my deepest gratitude to you for your constant encouragement and guidance. I have never met anyone who can always find a solution for any of my questions. You are so talented in so many fields of science. I could not thank you enough for taking me as a Ph.D. student of yours. Every time when I am lost in science you can always pull me back to the right track. Without your consistent and illuminating instruction, this thesis could not have reached its present form. I always enjoy working with you, it was a pleasant journey doing science with you for the past 4 years.

I would like to express my heartfelt gratitude to my colleagues at the Max Planck Institute for Chemical Ecology for all the help, and for both scientific discussions and friendship. I thank **Matthias Schöttner** for teaching me analytical chemistry and how to use analytical instruments; **Thomas Hahn** for solving all the instrument problems; **Mario Kallenbach** for your help with setting up instruments; **Klaus Gase** for helping me design primers. My big thanks would go to our Mensa team, you guys made my everyday lunch delightful. Thanks to **Youngsung Joo, Felipe Yon, Sven Heiling, Arne Weinhold, Michael Stitz, Youngjoo Oh, Martin Schäfer, Mariana Stanton, Variluska Fragoso, Matthias Schöttner, Van Thi Luu, Erica McGale, Meredith Schuman.** I enjoyed very much talking

to you during lunch and the afterwards coffee break. With this, I have learned a lot that I can not learn from work. My big thanks would next go to our Chinese community, **Zhihao Ling, Han Guo, Shuqing Xu, Wenwu Zhou, Ran Li, Jiancai Li, Dechang Cao, Ming Wang, Xiang Li,** you have made my stay in Jena feel like home. Thanks to all the scientific and non-scientific discussions in the lab, during lunch and coffee break. I would also thank my **basketball team** in Jena, you have made my leisure time spend to the fullest.

Many thanks to **Erica McGale** for proofreading and correcting my English; **Lucas Cortés Llorca** and **Sven Heiling** for helping me with silencing efficiency experiment; **Martin Schäfer** and **Thomas Brockmöller** for helping me with the Zusammenfassung in German. You are the people who helped me finish my thesis in time.

Last my thanks would go to my beloved **family** for their supports, they constantly encouraged me when I felt frustrated during my Ph.D..

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und unter Verwendung der angegebenen Hilfsmittel, persönlichen Mitteilungen und Quellen angefertigt habe.

Jena, den 08. 05. 2016 **Dapeng, Li**